Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *University of Michigan*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

"Fate" of First-List Associations in Transfer Theory: J. M. BARNES AND B. J. UNDERWOOD 97
Studies of Distributed Practice: XIX. The Influence of Intralist Similarity With Lists of of Low Meaningfulness: B. J. UNDERWOOD AND R. W. SCHULZ 106
Mixed Vs. Unmixed Lists in Transfer Studies: H. M. TWEDT AND B. J. UNDERWOOD 111
Individual Patterns of Physiological Activity as a Function of Task Differences and Degree of Arousal: M. M. SCHNORE 117
Changes in GSR to a Single Stimulus as a Result of Training on a Compound Stimulus: W. W. GRINGS AND T. N. SHMELEV 129
Effects of Threat of Shock, Distraction, and Task Design on Performance: R. E. MURPHY 134
Some Effects of Introducing and Withdrawing Knowledge of Results Early and Late in Practice: E. A. BILODEAU, I. MCD. BILODEAU, AND D. A. SCHUMSKY 142
Differential Effects of Continuous Extinction and Discrimination Training on the Generalization Gradient: W. K. HONIG, D. R. THOMAS, AND N. GUTTMAN 145
Time Estimation as a Function of Level of Behavior of Successive Tasks: R. K. S. DEWOLFE AND C. P. DUNCAN 153
One Determinant of Judged Semantic and Associative Connection Between Words: J. H. FLAVELL AND E. R. FLAVELL 159
The Effects of Visual Noise and Locus of Perturbation on Tracking Performance: W. C. HOWELL AND G. E. BRIGGS 166
Two-Choice Behavior of Rhesus Monkeys in a Noncontingent Situation: W. A. WILSON, JR. AND A. R. ROLLIN 174
The Role of Muscular Tension in the Recall of Interrupted Tasks: D. W. FORREST 181
Reaction Time as a Function of Foreperiod Duration and Variability: L. KARLIN 185

American Psychological Association

Vol. 58, No. 2 August 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 58, NO. 2 AUGUST, 1950

"FATE" OF FIRST-LIST ASSOCIATIONS IN TRANSFER THEORY

JEAN M. BARNES AND BENTON J. UNDERWOOD

Northwestern University

Theorists who attempt to arrive at an accounting of the facts of transfer and retroactive inhibition in verbal materials must inevitably make a decision concerning "what happens" to first-list associations when a second list is learned. The present study was designed to gather data which would make such decisions easier. Two transfer paradigms were used. One of these, A-B, A-C, is normally associated with production of negative transfer. In the other paradigm, A-B, A-B′, stimuli are identical with responses highly similar; such an arrangement normally produces positive transfer.

The data from the present study allow an evaluation of three different conceptions concerning the fate of first-list (hereafter called List 1) associations during the learning of the second list (hereafter called List 2). These three positions will be examined first as they apply to the A-B, A-C paradigm.

1. List 1 associations are unlearned or extinguished during the learning of the second list. While A-B, A-C is more comparable to the operations of counterconditioning than to extinction of conditioned responses, the term extinction will be used here in keeping with past practice. As a theoretical position, the unlearning or extinction hypothesis is associated with Melton and Irwin (1940) who used it to explain certain facts of RI. They suggested that at least part of RI could be accounted for by the loss of associative strength of List 1 responses resulting from extinction. It might also be assumed that negative transfer results from the interference accompanying the extinction process. A rather impressive amount of evidence could be marshalled in support of this extinction hypothesis. For present purposes, however, only an illustration of this evidence will be given. At various points in learning the second list, Briggs (1954) presented the stimuli common to both lists, one at a time, and asked *S* to give the first of the two responses which came to mind. The results showed that during the learning of List 2 there is a gradual decrease in frequency of List 1 responses, the curve being not unlike an extinction curve. A phenomenon analogous to spontaneous recovery of the List 1 responses was also found when retention was measured at various intervals of time after List 2 learning. Such evidence appears quite in line with the extinction hypothesis. However, as Briggs points out, a certain amount of ambiguity remains; the extinction curve does not necessarily mean that List 1 associations are extinguished in the sense that they are no longer available for recall. His results may mean only that List 2 associations become stronger than List 1 associations, hence, occur more and more frequently in recall. Such an increase in frequency of List 2 associations must necessarily be accompanied by a decrease in frequency of List 1 associations at recall. The question still remains,

therefore, as to whether or not the A-B association is available after A-C is learned and this is one question the present study attempts to answer.

2. The system of associations in List 1 (A-B) remains relatively independent and intact throughout the learning of the associations in List 2 (A-C). This will be referred to as the independence hypothesis. The verbal stimuli are, of course, identical in the A-B, A-C paradigm, but there are other possible differentiating cues (e.g., first list vs. second list) which would provide a means of having two independent response systems attached to the same apparent stimulus. Negative transfer could be said to occur in the process of establishing the independent response systems early in learning A-C. Retroactive inhibition could be attributed to loss of differentiation so that response competition occurs.

3. A third possible conception makes use of mediation. Having learned A-B in List 1, *S* retains this association so that List 2 items are learned with B as the mediator; *S* learns A—B—C. Negative transfer and RI could occur as a consequence of confusion on the part of *S* as to which response is used as the mediator and which response is mediated.

In the present study, different groups of *S*s were stopped at various points in learning List 2 and were asked to give *both* List 1 and List 2 responses to each stimulus. If (as degree of List 2 learning increases) there is an increasing inability to give List 1 responses (over and above normal forgetting), the extinction hypothesis would be favored. If there was no such loss either the independence hypothesis or the mediation hypothesis would be favored. Provisions were included in the design to choose between these latter two if it became necessary. For example, if the mediation hypothesis is tenable it should be found that B is given before C more frequently than C before B when *S* is asked to write down both responses. The independence hypothesis would not predict such an ordering in response recall.

Turning next to the paradigm in which stimuli are identical but responses highly similar (A-B, A-B′), the present study should again provide bases for choice among the three hypotheses.

1. Because retroactive facilitation is expected with this paradigm (e.g., Young, 1955) it might appear than an extinction hypothesis is not tenable. However, by the use of extinction plus other factors it is possible to account for retroactive facilitation. It is a fact that more List 1 instrusions occur in the learning of List 2 for this paradigm than for the A-B, A-C paradigm. Thus, the nonreinforcement of these responses makes extinction a very plausible mechanism. So, it could be assumed that in learning A-B′, A-B is extinguished. Retroactive facilitation could then be accounted for by mediation. It is known that highly similar items have high associative connection (Haagen, 1949). When *S* is asked to recall A-B after learning A-B′, even if A-B is extinguished the mediation sequence in recall could be A—B′—B, since there has been no extinction of the strong association between B′ and B. However, without the addition of still more factors the positive transfer in learning A—B′ cannot be accounted for. Since the results of the present procedures do not make the extinction hypothesis a reasonable one for this paradigm, no attempt will be made here to suggest the additional factors needed to account for the positive transfer produced by this paradigm.

2. The independence hypothesis, used in conjunction with a theory of response generalization (Underwood, 1951), can be employed to account for both positive transfer and retroactive facilitation in the A-B, A-B′ paradigm.

3. The mediation hypothesis also provides an attractive alternative for this positive transfer paradigm. As noted above, items which have high similarity have high associative connection. Thus, having learned A-B, the learning of A-B-B′ should be very simple since B and B′ are already strongly connected. Therefore, positive transfer should occur. In learning List 2, if mediation occurs *S* will be given additional practice on A-B; hence, this association will be further strengthened during List 2 learning, and retroactive facilitation would be expected to occur.

In summary, two paradigms are used, one associated with the production of negative transfer, the other

with production of positive transfer. By requiring *S* to recall both List 1 and List 2 responses at various points in the learning of List 2, it was expected that choice could be made among three conceptions (extinction, independence, mediation) of the fate of List 1 associations during the learning of List 2.

METHOD

General.—Two parallel experiments were conducted, one using the A-B, A-C transfer paradigm and one using the A-B, A-B′ paradigm. Within each experiment there were four groups of *Ss*. All *Ss* learned a first paired-associate list (List 1) of eight pairs to one perfect trial. Then List 2 was presented for a specified number of trials, namely 1, 5, 10, and 20 anticipation trials for the four groups in each experiment. After the specified number of trials for a given group, the memory drum was stopped, *S* was provided a sheet of paper on which the eight stimuli were printed, and he was asked to write down the two responses (one from the first list and one from the second) which were associated with each stimulus. This written recall provided the major source of data.

Lists.—Eight nonsense syllables of from 60% to 73 % association value (Glaze, 1928) were used as stimuli for all lists. Intra-stimulus similarity was low in that no consonant was repeated, and four vowels were used twice each. The responses were two-syllable adjectives taken from Haagen (1949). In the A-B, A-C lists there was no apparent interlist response similarity and intralist response similarity was as low as careful inspection procedure will produce. In the A-B, A-B′ lists the responses again had low intralist response similarity but high interlist response similarity. The responses from each list which were paired with the same stimulus had similarity ratings ranging from .9 to 1.4 by Haagen's scale. These high-similarity response pairs were as follows: *insane—crazy; barren—fruitless; complete—entire; royal—regal; double—twofold; afraid—scared; tranquil—peaceful; spoken—verbal.* Three different pairings of stimuli and responses were used for both experiments to avoid the possibility that fortuitous associations between stimulus and response would bias the results. An equal number of *Ss* in each group was given each pairing. For each experiment for each condition a given list was used as List 1 for half the *Ss* and as List 2 for the other half. Four orders of items were used to minimize serial learning.

The lists were presented on a Hull-type drum at a 2:2-sec. rate. The intertrial interval was 4 sec. Anticipation learning was used for List 1 and for List 2 to the point at which written recall was given (after 1, 5, 10, or 20 anticipation trials). A 1-min. interval separated the learning of the two lists.

Subjects.—A total of 192 *Ss* was used. There were 24 *Ss* in each condition for each experiment, the conditions being the point at which *S* was asked for written recall. All *Ss* had previously served in at least one verbal-learning experiment. However, no other experiment had been run in which *S* was asked to recall responses from both lists, so there is no reason to believe that *S* anticipated that he would be asked for such recall. It should be emphasized that each *S* was asked for written recall only once so that it is not believed that *Ss* made any special attempt to remember List 1 responses. In assigning *Ss*, a listing of 192 entries was made such that each of the eight conditions occurred 24 times. Within these limits the ordering was random and *Ss* were successively assigned in order of their appearance at the laboratory.

Written recall.—When the specified number of trials on List 2 had been given, the memory drum was stopped and the sheet for written recall given to *S*. The stimuli were listed on the sheet with two blank spaces under each. The *Ss* were first instructed to write down the responses from the two lists below the appropriate stimuli. They were further told to write these down as they came to mind, and not to attempt to recall all the responses from one list first and then all those from the other. Two minutes were then allowed for this initial recall. Following this, they were instructed to go through the responses they had written, and indicate, by assigning the numbers 1 and 2, from which list the responses had come. An additional 2 min. were allowed for this. The *Ss* were also told to write down any additional responses which they might think of while they were assigning list numbers.

After the recall papers were collected, a series of questions was asked to determine: (*a*) if *S* had written the responses as they came to mind or if there was deliberate attempt to recall the adjectives from a given list first, (*b*) whether or not *S* had used first-list responses to mediate the learning of the second-list responses.

Results

Transfer.—The major concern in the present paper is with the written recall results. However, it will be worthwhile to note briefly the transfer facts and to show that random assignment of *Ss* to groups was effective. For all eight groups combined, the mean number of trials to learn List 1 to one perfect trial was 10.36, with the means for groups ranging from 9.17 to 12.00 trials. The F was less than 1. Lacking a precise control for transfer, some indication of the effects can be obtained by comparing the learning of List 1 with that of List 2. The total correct responses for the first three trials of the two lists were used as the response measure for the three groups having 5, 10, and 20 trials on List 2. For the A-B, A-B′ paradigm, more correct responses were given on List 2 than on List 1, indicating positive transfer. For all three groups combined, the mean number correct on the first list was 9.84, and for the second, 14.35. The t was 8.55. For A-B, A-C the mean number correct on the first list was 9.64, and on the second, 8.57. This t was 1.67. This result indicates that in a statistical sense no negative transfer was present, but this is not an unusual finding for this type of comparison since learning-to-learn and warm-up would counteract the negative effect. Finally, in keeping with previous studies (e.g., Underwood, 1951), the number of overt intrusions of List 1 responses during the learning of List 2 was much greater for A-B, A-B′ than for A-B, A-C.

Written recall.—There were several alternative ways by which the written recall could be scored. The most stringent method would be to count a response as correct only if it were placed with the appropriate stimulus and also correctly identified as to list. The most liberal method would be to count an adjective correct if recalled, regardless of whether it was placed with the correct stimulus and correctly identified as to list. The results for both methods are shown in Table 1. An inspection of this table will show that the differences produced by the two scoring methods are relatively small. The implication is that if *S* wrote down a response, he usually identified it correctly with stimulus and list.

TABLE 1

Mean Number of Responses Recalled From Each List for Each Transfer Paradigm as Scored by Two Methods[a]

Method	List	Point of Written Recall (Trials)			
		1	5	10	20
Paradigm A-B, A-C					
1	1	6.67	5.38	5.00	4.12
	2	3.46	6.29	7.33	7.38
2	1	6.96	5.58	5.42	4.29
	2	4.12	6.71	7.71	7.42
Paradigm A-B, A-B′					
1	1	7.67	7.21	7.12	6.92
	2	7.21	7.25	7.83	7.79
2	1	7.75	7.33	7.25	7.08
	2	7.38	7.29	7.96	7.83

[a] In Method 1 the response is counted correct only if correctly identified with stimulus and list. In Method 2 it is counted correct if merely recalled.

Turning first to the results for the A-B, A-C paradigm, the results for the stringent scoring method have been plotted in Fig. 1. As would be expected, as the number of learning trials on List 2 increases, the number of correct responses given from this list increases. Responses from the first or A-B list, however, show a gradual decline as the number of trials on List 2 increases. It is as if

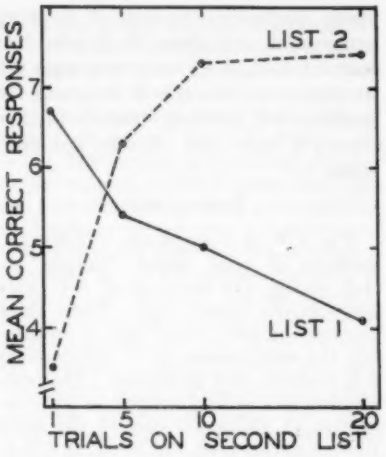

FIG. 1. Mean number of responses correctly recalled and identified with stimulus and list in the A-B, A-C paradigm.

the A-B associations are weakened or extinguished during the learning of A-C. The loss as a function of number of A-C trials cannot be attributed to a few *S*s losing many responses since most *S*s showed the decline. An analysis of variance of the four recall points for A-B gave an F of 8.86 ($F = 4.88$ for $P = .01$ with 3 and 92 *df*).

It was noted that Fig. 1 suggests that the A-B associations are extinguished. It might be suggested, furthermore, that not only are these associations extinguished, but also that the response per se is not available. This could be deduced from the fact that the two scoring methods did not differ much, implying that if *S* knew the response, he also knew what stimulus and what list it went with. However, this conclusion cannot be reached with complete confidence. The reason is that the instructions to *S* stressed the recall of the responses to the specific stimuli; he was never told to put down a response even if he didn't know with which stimulus it was paired. It is therefore possible that *S*s did not put down responses unless they were reasonably confident of its stimulus.

The objection might be raised that the decline in A-B associations seen in Fig. 1 is not to be attributed to the learning of A-C but rather represents simple forgetting. To make sure that this would not be a valid objection, a control group of 12 *S*s was run. This group learned List 1 to one perfect trial and then rested for a period of time equivalent to that spent in learning A-C by the group given 20 trials. This period was 13 min. (including the 1 min. between lists). The rest interval was filled with *S* working on a pyramid puzzle. After the rest, *S*s were given a written recall for the responses to the stimuli in the list they had learned. Counting only those responses which were appropriately paired with their stimuli, the mean recall was 7.75. Hence, it can be concluded that the decrement in recall of A-B associations as a function of trials on A-C cannot be attributed to simple forgetting, but must result from the learning of A-C.

The results for the A-B, A-B′ paradigm as given in Table 1 are quite different from those of A-B, A-C. It may first be noted that the recall of List 2 responses (A-B′) is nearly perfect after only one anticipation trial and, of course, remains high throughout 20 trials. Indeed, by both scoring methods the median recall for all groups at all points is 8.0 items. The first-list associations (A-B) show no appreciable decline. By the stringent scoring method the mean recall, even after 20 trials on A-B′, is nearly seven.

Order of recall.—The facts concerning the order of recall of the two

responses can be used as evidence for or against a mediation hypothesis, as this hypothesis was discussed in the introduction. If *S* tends to recall List 1 responses first, it could be taken as evidence in support of a mediation hypothesis. Contrariwise, it would be evidence against mediation if List 2 responses tend to be recalled first. In this analysis data were not used from 12 *S*s who admitted under questioning that they attempted to write down all responses from one list first and then all from the other. Of these, seven had learned A-B, A-C, and five, A-B, A-B′. Obviously, only cases in which both responses were recalled could be used. For A-B, A-B′, there were 161, 159, 170, and 162 pairs of responses recalled for 1, 5, 10, and 20 trials, respectively. The corresponding percentages for the List 1 response being recalled before the List 2 response are 79%, 62%, 58%, and 53%. For A-B, A-C, the number of cases where both responses were recalled was 79, 113, 109, and 95 for 1, 5, 10, and 20 trials, respectively. The percentages of List 1 responses recalled first are 81%, 49%, 40%, and 43%. Assuming chance order as 50%, it can be seen that (excepting the conditions where only one trial was given on List 2 and where List 1 would be much stronger than List 2) with A-B, A-C the List 1 responses tend to be recalled first with chance or less than chance frequency. For A-B, A-B′ first-list responses are recalled with greater than chance frequency although after 20 trials this frequency has nearly reach 50%.

Subject reports.—Of the 96 *S*s who learned the A-B, A-B′ paradigm, 94 said they had used the List 1 responses to mediate the learning of List 2. These *S*s further reported that they tended to drop the use of these mediators as List 2 learning progressed. Of those 96 *S*s who had learned A-B, A-C, only two reported attempts to use the B response to mediate A-C learning; both *S*s further reported that this merely confused them.

Discussion

The A-B, A-C paradigm.—Three conceptions of what might "happen" to A-B during the learning of A-C were given in the introduction. These three conceptions were extinction or unlearning, the maintenance of two independent S-R systems, and mediation. The present data give strong support to the conception of extinction and at the same time argue against the other two conceptions. The recall of List 1 associations decreased progressively throughout the learning of A-C with nearly a 50% loss after 20 trials on A-C. A hyperbolic equation fitted to this curve predicts an asymptote at 3.46. Thus, it does not seem that all items would be extinguished, even with an extremely large number of trials on A-C. This conforms to the fact that forgetting in the RI paradigm has not been shown to be complete with very high degrees of interpolated learning. Thus, while the present data strongly support an extinction hypothesis they do not indicate why all items are not, nor are likely not to be, extinguished.

It seems reasonable to reject the independence hypothesis. If this hypothesis is to be retained, some mechanism will have to be added to account for the loss in the A-B associations over and above normal forgetting. If it is said that the A-C associations were so dominant that *S* was unable to recall the A-B associations, then the conception becomes very similar to an extinction hypothesis.

The mediation hypothesis cannot be seriously considered in light of the evidence. If List 2 responses are learned through mediation of B (A—B—C) the A-B association should not be lost from *S*'s repertoire. If it is said that the

mediators drop out, then something like extinction must also be added to account for their loss from *S*'s repertoire. Furthermore, *Ss*' reports did not indicate that mediation had taken place (as was true of the reports for A-B, A-B′). All in all, the results give strong support to an extinction-like process; they do not give much support to an hypothesis of independence nor one of mediation.

It was noted that if *S* could recall a response he could usually identify it correctly as to list. On the surface, such evidence seems contrary to conceptions which say that some RI may be attributed to competition resulting from a lack of differentiation as to which response belongs in which list (e.g., Underwood, 1949). However, two additional facts must be noted which are relevant to this matter. First, in the present study essentially unlimited recall time was given so that identification of responses with lists should be more accurate than when recall is paced by a memory drum at a 2-sec. rate. Secondly, this differentiation is assumed to be very high when recall is given immediately after interpolation; differentiation is assumed to decrease as a function of time between interpolated learning and recall (Underwood, 1949). It is quite possible that the high accuracy of identifying responses with lists would be lost even with the present method if, say, a 24-hr. interval occurred between interpolation and written recall. Nevertheless, the present results would strongly suggest that nearly all RI could be accounted for by unlearning or extinction if unlimited recall time is given immediately after interpolated learning. The lack of differentiation, leading to competition between responses, should become an additional component in RI only after the passage of time since interpolated learning. Indeed, in view of the findings suggesting spontaneous recovery of extinguished associations (Briggs, 1954), it can be held that after the recovery is complete, all RI is due to competition. This two-component conception of RI will incorporate the major facts.

The A-B, A-B′ paradigm.—The data for this paradigm have shown that the A-B associations were maintained at a high level throughout the learning of A-B′. The very slight decline noted over 20 trials could be attributed to the extinction of the A-B responses for those cases in which *Ss* did not know the meaning of a word or words, hence, similarity of the responses was ineffective and the results were the same as an A-B, A-C paradigm. However, while accepting this possibility, it can be omitted from further consideration. The question still remains as to whether it is or is not likely that the A-B associations were extinguished while *S* learned A-B′ even when similarity was effective. It was noted in the introduction that such an hypothesis was a plausible one; indeed, looking only at the basic results it remains plausible.

The possibility that A-B could be extinguished and the present results still be obtained (no apparent loss in the A-B associations) may be handled by mediation in which after learning A-B′ (and extinguishing A-B), *S* recalled the B response because of its high associative connection with B′. Thus, when asked to recall, the associations were A to B′ and then B′ to B. If this took place, A-B could be extinguished and the present basic results would still be obtained. However, other findings argue against this interpretation and tend instead to support mediation in which A leads to B and B leads to B′ at recall.

1. If B is recalled via B′, B′ should be recalled before B. The data indicate that B is more often recalled first. It is true that after 20 trials B and B′ were recalled first about equally often. So, it is still possible that with high degrees of A-B′ learning, some extinction of A-B did occur.

2. Nearly all *Ss* report using B as a mediator to learn A-B′; none reported that B′ was used as a mediator in the recall of B. However, they did report that the use of B as a mediator tended to drop out as learning of A-B′ proceeded. Some extinction of A-B might have been a concomitant or cause of this drop out. But, the fact that the percentage of times in which A-B was recalled first (before A-B) never drops below 50%

suggests that if extinction was occurring it was much less in amount than that which occurred in A-B, A-C.

The conclusion is, therefore, that while the possibility of extinction of A-B in the A-B′ paradigm cannot be completely ruled out, it does not appear as compelling an interpretation as does the mediation hypothesis in which A-B mediates B′.

Finally, what about the hypothesis of independent response systems in which response generalization is used to account for positive transfer and retroactive facilitation? The use of response generalization to explain certain facts of transfer has been of considerable help in organizing the facts of transfer. One particular statement of this formulation can be briefly summarized (Underwood, 1951). It is assumed that when A-B is being learned, all responses similar to B likewise develop some associative strength to A. The amount of such strength, developed through what has been called parasitic reinforcement, is directly related to similarity of the response with the B response. Positive transfer occurs, therefore, because less learning has to occur in the second list than in the first. Thus, in the present situation, if A-B′ develops some associative strength as a consequence of learning A-B, positive transfer should occur when *S* is asked to learn A-B′. Retroactive facilitation would occur, since during the learning of A-B′ the A-B association would be further strengthened. It should be noted that this hypothesis deals only with the associations between A and B, and between A and B′; it does not deal in any way with the already well-established association between B and B′ which forms a central part of the mediation hypothesis.

The present evidence suggests that the response-generalization accounting of certain facts of transfer may well be abandoned in favor of a mediation hypothesis. The mediation hypothesis will account for the same general facts as will response generalization; furthermore, certain facts of the present experiment seem difficult to reconcile with the response-generalization hypothesis.

1. Almost without exception, *S*s report mediation. While it may be possible to extend the theory of response generalization to account for this fact that *S*s report mediation, the mediation per se can be used to account for the results, without further elaboration.

2. It was noted that after only one anticipation trial, recall of List 2 was nearly perfect. The theory of response generalization assumes a generalization gradient decreasing rather sharply as similarity decreases. The generalized association is not assumed to grow in associative strength at a rate comparable at all to the directly-reinforced association (A-B). To account for the extraordinarily rapid learning of A-B′ would mean that there is essentially no gradient between B and B′, or to say this another way, the gradient is flat. In previous experiments, no such rapid learning of A-B′ was noted; learning was more rapid than for a control condition but nothing like that shown here in the written recall. Indeed, in the present study for the groups having 5, 10, and 20 trials (before written recall), the mean number of items recalled on the second anticipation trial (comparable to the point at which written recall was asked for in the group having only 1 anticipation trial) was 5.04. This is to be contrasted with 7.75, obtained on written recall at the same point. The reason for this discrepancy, it is believed, is that under the standard procedures of anticipation learning *S* does not have time to mediate all items. Given more time, recall is nearly perfect right at the start of "learning" the second list. This finding is not compatible with the theory of response generalization.

The conclusion is that mediation of B′ through A-B is a more appropriate formulation than response generalization to account for positive effects in transfer and retention when response similarity is involved with identical stimuli. As similarity between responses decreases, the associative connection between them will likewise decrease; thus, a gradient-like phenomenon will appear in transfer. With moderate similarity between responses, the associative connection between the two will be less than in the present case, and additional strengthen-

ing of the associative connection will be necessary before mediation is completely effective. Thus, positive transfer will be less than when the responses are highly similar. It is apparent that since no evidence for mediation could be found in A-B, A-C, the mediation hypothesis must be abandoned when the responses reach a certain level of dissimilarity (or a low level of similarity). It is at this point that extinction begins to occur with consequent negative transfer.

Summary

The A-B, A-C and the A-B, A-B′ transfer paradigms were studied in order to evaluate three conceptions concerning the fate of first-list associations in learning a second list; namely: extinction of first-list associations, maintenance of two independent S-R systems, and mediation of the learning of the second-list response by the first-list association. For each paradigm, different groups of 24 *Ss* each were stopped after 1, 5, 10, or 20 anticipation trials on List 2 and were asked to write down *both* List 1 and List 2 responses to the stimuli. Both lists contained eight pairs of nonsense syllables and two-syllable adjectives. List 1 learning was carried to one perfect trial.

For the A-B, A-C paradigm there was a gradual reduction in reproduction of List 1 responses as degree of List 2 learning increased. A control condition showed that this reduction could not be accounted for by normal forgetting. No successful mediation of the C response via A-B was reported. It was concluded that of the three alternative conceptions, extinction of the List 1 responses was clearly to be preferred. It appears that nearly all retroactive inhibition measured immediately after interpolated learning may be due to extinction or unlearning of first-list responses during the learning of the second list.

For the A-B, A-B′ paradigm, the List 1 responses showed no appreciable loss over 20 trials and List 2 was given nearly perfectly after one anticipation trial. Of 96 *Ss*, 94 reported the use of the A-B association to mediate the B′ response. The rapid learning of List 2 is understandable by the mediation hypothesis, but some extinction of the A-B association in the A-B, A-B′ paradigm cannot be ruled out completely by the present data. The A-B association may be extinguished and then mediation of the List 1 response occurs via A-B′-B. However, considering all evidence, it seemed most probable that mediation occurs most frequently in the order A-B-B′. Finally, the evidence suggests that transfer effects produced by variation in response similarity can be more simply accounted for by mediation than by a theory using response generalization.

References

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Haagen, C. H. Synonymity, vividness, familiarity, and association-value ratings for 400 pairs of common adjectives. *J. Psychol.*, 1949, **30**, 185–200.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 175–203.

Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

Young, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, **50**, 113–119.

(Received August 18, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

STUDIES OF DISTRIBUTED PRACTICE: XIX. THE INFLUENCE OF INTRALIST SIMILARITY WITH LISTS OF LOW MEANINGFULNESS[1]

BENTON J. UNDERWOOD AND RUDOLPH W. SCHULZ

Northwestern University

In the immediately preceding study by Underwood and Richardson (1958) the interactions between meaningfulness and intertrial interval and between intralist similarity and intertrial interval were studied. Massed practice was defined as 8 sec. between trials, distributed practice by 38 sec. The findings in this study, which form the immediate background for the present one, follow.

1. With a serial list constructed of syllables of from 93–100% association value (Glaze, 1928) and having low intralist similarity, distributed practice and massed practice produced equivalent rates of learning. With a list of syllables of the same association value but with *high* intralist similarity, distributed practice was markedly superior to massed practice. Thus, these results indicated a strong interaction between intertrial interval and intralist similarity; the greater the intralist similarity the greater the effect of distributed practice. The generality of this finding, however, was questioned by the second major finding.

2. With serial lists constructed of syllables of from 0–20% association value and of low and high intralist similarity, the results were inconclusive. While distributed practice resulted in somewhat better performance than did massed practice for both levels of intralist similarity, the differences were not significant and there was no evidence of interaction between intralist similarity and intertrial interval as was found with the lists of high meaningfulness.

These inconclusive results with the lists of low meaningfulness prompted the present investigation in which only lists of low meaningfulness are used. Several considerations directed the choice of conditions. (*a*) In the previous study a total of 30 acquisition trials was given under all conditions. The level of learning for the high-meaningful lists was considerably higher than for the lists of low meaningfulness. It is possible that with additional trials the expected interaction between intertrial interval and intralist similarity would emerge for the lists of low meaningfulness as it did for the lists of high meaningfulness. Therefore, 40 acquisition trials were given in the present experiment. (*b*) The effect on learning of a given difference in formal intralist similarity seems to be greater with lists of low meaningfulness than for lists of high meaningfulness. Why this may be true was discussed in the previous paper, but just what implication this has for a possible interaction between intertrial interval and intralist similarity for lists of low meaningfulness is not clear. In any event, it seemed necessary to make sure that a broad range of intralist similarity was used. Consequently, the range of formal intralist similarity has been extended in the present study. (*c*) Two of the four lists used are the same as those used in the previous study. In order to explore more fully the effect of length of intertrial interval, the two intervals used here are 2 sec. and 17 sec. (as opposed to 8 sec. and 38 sec. in the earlier study). The learning of two lists, therefore, can be evaluated as a function of four different intertrial intervals.

[1] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

METHOD

Lists.—Four serial lists were used, each consisting of 10 nonsense syllables having a Glaze (1928) association value of from 0–20%. The four lists were as follows:

List 1	*List 2*	*List 3*	*List 4*
CEF	SIW	GOK	XAZ
DAX	QAS	GUW	ZAX
GAH	QUW	GYK	XEY
MEQ	KOJ	KIG	YOZ
SIJ	BEJ	KEZ	ZUY
TOV	YAB	WUG	YUZ
NUB	YOV	WEZ	XIY
YIL	XEZ	ZIK	ZEY
ZOK	ZIK	ZYW	YIX
WUP	VUX	ZOW	ZOY

Intralist similarity increases from List 1 through List 4 as follows: List 1 has no duplicated consonants; in List 2, 10 consonants are used twice each; in List 3, four consonants are used five times each, and in List 4, one consonant was used eight times, one seven, and one five times. A vowel occurs twice in each list. Repeated consonants were not used systematically as first or as last letters, but occurred in both positions. Lists 1 and 3 were used in the previous study (Underwood & Richardson, 1958).

Conditions and Ss.—Each list was learned by two groups of *Ss*, one having a 2-sec. intertrial interval and one having a 17-sec. intertrial interval. Thus a total of eight groups of *Ss* was used and each group contained 30 *Ss*, all naive to verbal learning in the laboratory. A randomized order of lists and intervals was made up subject to the limitation that each list occur 30 times with each interval. The *Ss* were assigned successively in order of their appearance at the laboratory.

Ten different serial orders of the items in each list were used such that each item occurred once at each serial position. Three *Ss* in each group learned with a given order. Three horizontal asterisks were used as the anticipatory cue for the first syllables. Rate of memory-drum presentation was 2-sec. per item with the anticipation method being used throughout. A total of 40 such anticipation trials was given. The groups having distributed practice worked on a cancellation task during 15 sec. of the 17-sec. distribution interval.

RESULTS

Total correct over 40 trials.—The mean total correct responses over 40 trials are shown in Fig. 1. The baseline is plotted as four equally separated points although no assumption is made that the successive lists represent equal increments in intralist similarity. It is to be noted first that under massed learning performance decreases as intralist similarity increases. The second fact shown in Fig. 1 is that distributed practice consistently leads to better performance than does massed practice. Both of these above facts are substantiated statistically. The F for Lists (Intralist similarity) is 15.27; with 3 and 232 df, the F at the 1% level is 3.88. The F for intertrial interval is 29.86; with 1 and 232 df, the F at the 1% level is 6.76. There is no statistical evidence of heterogeneity of variance. From Fig. 1 it also appears that the effect of distribution is greater the greater the intralist similarity. If such were true, it would substantiate the finding of the earlier study with lists of high meaningfulness. Statistically speaking, this interaction falls just short of the 5% level. The F is 2.58; with 3 and 232 df, an F of 2.65 is needed for significance at the 5% level. Therefore, on this evidence alone, there is no strong support for the

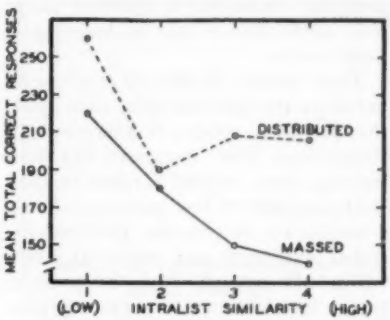

FIG. 1. Total correct responses over 40 trials as a function of intralist similarity and intertrial interval.

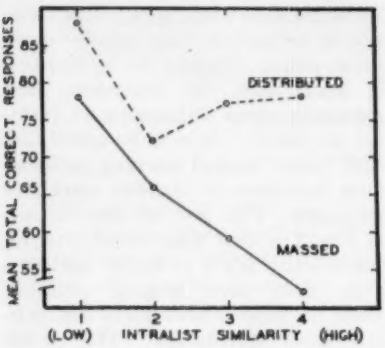

FIG. 2. Total correct responses on Trials 31–40 as a function of intralist similarity and intertrial interval.

interaction between intertrial interval and intralist similarity.

It will be noted that the difference in performance resulting from massing and distribution is greater for List 1 than for List 2. It is not known whether this results from some peculiarity in a list or whether a sampling variation is involved. Complete learning curves show that the difference in massing and distribution for List 1 is evident throughout most of the 40 learning trials, whereas for List 2 the curves clearly separate only after 20 trials. As a second point of analysis, therefore, a measure using only performance late in learning is used.

Total correct on last 10 trials.—As noted in the introduction, data from the previous study (Underwood & Richardson, 1958) suggested that had learning been carried further in that study for lists of low meaningfulness, a significant interaction between intralist similarity and intertrial interval would emerge. In the present study, 10 additional trials were given. On the assumption that the effects of massing may be accumulative, a test for the interaction has been made over the last 10 trials in the present experiment. The mean total correct responses on Trials 31–40 are shown in Fig. 2. Again, the effects of the two major variables (intralist similarity and intertrial interval) are highly significant. Furthermore, the interaction between the two variables gives an F of 3.33, with an F of 2.65 needed for the 5% level, and 3.88 for the 1% level. The interpretation is not completely unambiguous, however, since the Hartley test shows the maximum difference in variances to be significant beyond the 1% level. Nevertheless, it does not seem unreasonable to conclude that there is a true interaction effect. It will be noted that the effect of distributed practice with Lists 3 and 4 is large; the performance on these lists is approximately the same as that for List 1 (low similarity) learned under massed practice. Thus, distributed practice almost completely cancels the deleterious effects of increasing intralist similarity.

Co-ordination with previous experiment.—In the previous experiment (Underwood & Richardson, 1958) Lists 1 and 3 were used with intertrial intervals of 8 sec. and 38 sec. The *Ss* for both experiments were drawn from the elementary psychology course and all were naive to verbal learning. No large differences in ability level between the groups of *Ss* in the two experiments would be expected. In the previous experiment 30 anticipation trials were given; the data for the two experiments are brought together in terms of mean number correct on Trials 26–30. In the previous study 50 *Ss* were used for each condition. To equate N for the two experiments for statistical purposes, 30 of the 50 *Ss* were selected randomly subject to the restriction that each of the 10

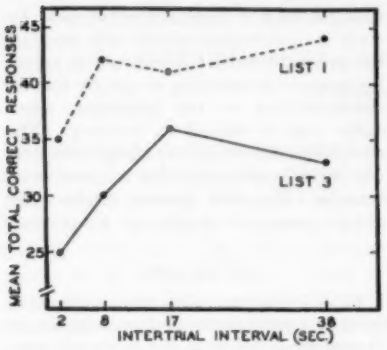

FIG. 3. Total correct responses on Trials 26–30 as a function of intralist similarity and intertrial interval. List 1 had relatively low intralist similarity, List 3, relatively high. The filled circles represent data from a previous experiment.

orders of the items for a given list was represented by three *S*s. The results are shown in Fig. 3. Again, both intertrial interval and intralist similarity are highly significant. The interaction term, however, falls far short of significance even though the change in performance between 2 sec. and 17 sec. for List 3 is much greater than for List 1. Beyond a 17-sec. intertrial interval no appreciable increase is produced by distributed practice.

Other findings.—Although no data on the matter were presented in the previous report, it was observed that the bowed serial-position curves were different for massing and distribution when distribution facilitated learning. Distribution facilitated learning more for items in the later serial positions than in the early positions. This effect is maximized in the present study for Lists 3 and 4 where the effect of distribution was the greatest. This may be seen in Fig. 4 where the mean number of correct responses for each serial position for Lists 3 and 4 combined are plotted. As can be seen, there is a gradual increase in the difference between the two curves for successive serial positions.

In all four lists in the previous study performance under the distributed conditions was poorer than under the massed for approximately the first six trials. At about this point a "switch-over" occurred and the distributed learning became superior. It was believed that this effect was due to one of two causes. (*a*) In distributed learning *S* had to learn to turn to the cancellation task at a signal at the end of a trial and to turn back to learning at a signal at the end of the cancellation period. Some confusion may have resulted from this procedure so that the learning was initially lower under the distributed condition than under the massed. (*b*) Another possible cause was that distributed practice allowed for a slower warming up than did massed practice. That is, if warm-up is lost during the distribution intervals, maximum warm-up would be attained more quickly by *S*s learning under massed conditions than by those learning under distributed conditions. In the present experiment there was absolutely no evidence for the switch-over effect. This would indicate that the second of the two above alternatives is a more plausible explanation than the first for the switch-over effect observed in the previous study. In the present study the *S*s serving under distributed-practice conditions would have had to learn to switch back and forth from learning to cancellation just as did the *S*s in the previous experiment. But, in the previous experiment, the distribution interval was 38 sec. as opposed to the 17 sec. used in the present study. It is therefore possible that loss of warm-up would be greater between trials with the longer distribution interval thus producing the switch-over effect which did not occur here with a shorter distribution interval.

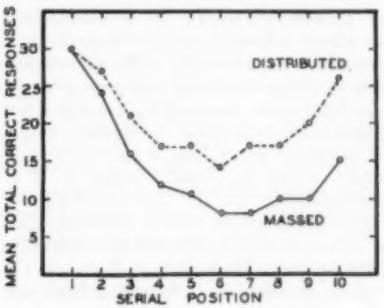

FIG. 4. Total correct responses over 40 trials as a function of serial position and intertrial interval. The data are for Lists 3 and 4 combined.

Discussion

The results clearly show that distributed practice facilitates the learning of lists of low meaningfulness. And, although there is some statistical ambiguity, the present results taken in conjunction with the ones from Underwood and Richardson (1958) indicate that intralist similarity is the critical interacting variable. The evidence points rather persistently to the fact that without a certain minimum degree of interference, distributed practice will not facilitate learning. In the list with the lowest degree of intralist similarity used here, there still remain duplicated vowels which could produce this minimum degree of interference necessary for distributed practice to facilitate learning. It is also possible that as meaningfulness gets lower and lower a new source of interference develops, namely, an interference from previously well-established letter-sequence habits. Whatever the source, it appears that a necessary condition for facilitation by distribution is some minimum number of conflicting response tendencies.

If one chose to say that some form of inhibition develops as a function of conflicting response tendencies and that this inhibition dissipates with relative rest, the present data would offer no serious problems. Such an inhibition would need to dissipate very rapidly for the following reasons: (*a*) The effects of distribution appear to be asymptotic at about 15 sec. even with very high intralist similarity; (*b*) The effects are less in the early serial positions than in the later ones, suggesting that even the 2 sec. between trials in the massed conditions allows some loss of inhibition but which builds up rapidly again as *S* is presented the list; (*c*) There is an appreciable difference in rate of learning with a 2-sec. intertrial interval as compared with an 8-sec. interval.

It is possible to construct a build-up function and a dissipating function for such an inhibition which will mediate the present results. The problem arises, however, in attempting to specify further characteristics of the inhibition, especially how it related to learning. The classical performance-type inhibition does not seem appropriate for a number of reasons. But, the present authors are not yet prepared to offer an alternative.

Summary

In an attempt to clarify some ambiguities found in a previous study, four serial lists of 10 nonsense syllables of low meaningfulness, varying in intralist similarity, were learned with a 2-sec. intertrial interval and with a 17-sec. intertrial interval. The eight groups of 30 *Ss* were naive to verbal learning. Each list was presented for 40 trials.

The results show that distributed practice facilitates learning at all levels of intralist similarity. While the statistical evidence on the matter is not entirely convincing, it was concluded that distributed practice produces a greater facilitation in lists of high intralist similarity than in lists of low intralist similarity. The effect of distribution is greater for items in the late serial positions than for those in the early positions. In comparing the results with those of a previous study it was concluded that intertrial intervals beyond about 15 sec. would not produce additional facilitation in learning.

It was pointed out that distributed practice facilitates learning only when some minimum number of conflicting responses are present. The results are compatible with an inhibition theory when this inhibition is produced by interference.

References

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Underwood, B. J., & Richardson, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56**, 213–219.

(Received August 14, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

MIXED VS. UNMIXED LISTS IN TRANSFER STUDIES

HELEN M. TWEDT AND BENTON J. UNDERWOOD[1]

Northwestern University

In relatively recent years there has been an increase in the frequency with which mixed lists (ML) have been used to study transfer with verbal materials. In ML a single group of *S*s may be used. Each *S* learns only two lists in which different relationships between subgroups of items allow the simultaneous observation of the transfer effects produced by these different relationships. Thus, if each of two lists consisted of nine paired associates, three pairs in each list could form the A-B, A-C paradigm, three the A-B, C-B, and three an A-B, C-D paradigm.

The ML procedure may be contrasted with the classical one in which unmixed lists (UL) are used. With UL the relationships between all items in the two lists are the same. Thus, if the paradigm was A-B, A-C, all stimuli in the two lists would be the same and all responses different. Obviously, to study transfer effects for more than one paradigm requires as many different sets of lists as there are paradigms.

Certain contradictions in the facts of transfer could be attributed to the possibility that UL and ML produce different transfer effects. The clearest example of this possibility is seen in the work of Mandler and Heinemann (1956) and Porter and Duncan (1953). In both of these studies an A-B, A-C paradigm was used, and also a paradigm that will be called A-B, A-Br. In this latter paradigm the same stimuli and same responses are used in both lists (UL) or among a subgroup of items in the two lists (ML), but they are re-paired in the second list. Porter and Duncan, using UL, found poorer second-list performance on A-B, A-Br than on A-B, A-C, while Mandler and Heinemann found just the opposite. There are other differences between the two studies over and above UL vs. ML which could account for this discrepancy. Indeed, a study (1958) published after the present one was completed essentially confirmed the Porter-Duncan results using ML. Therefore, other differences between the studies may be responsible for the different results. In other contradictions in the literature (e.g., Bugelski and Cadwallader [1956] vs. Osgood [1946]) ML vs. UL is again one of several possible factors involved. Therefore, a direct test of the transfer effects for the two kinds of lists is indicated to decide whether or not ML vs. UL is a variable of importance in the study of transfer. The major purpose of the present study is to make such a test.

There was an a priori reason to suggest that ML vs. UL is a variable influencing the relative amount of transfer among different paradigms. This may be seen by considering the situation in which *S* learns A-B, A-C with UL. While this paradigm can produce negative transfer, it has always been puzzling why (*a*) the negative effects are small, and why (*b*) so few B responses intrude during the learning of A-C. This would be less puzzling if it is assumed that two relatively independent response systems are maintained, differentiated

[1] We are grateful to R. W. Schulz for his critical reading of the manuscript.

by temporal or first and second list distinctions. If this differentiation is possible, *S* could instruct himself *not* to give responses from the first list and a relatively small amount of negative transfer would be expected. However, if the ML procedure is used in which some of the first-list responses are appropriate in the second list (A-B, C-B or A-B, A-Br) and some are not (A-B, A-C), the maintenance of the independent response systems is not possible. The prediction would be, therefore, that negative transfer would be greater for ML than for UL in the A-B, A-C paradigm. It is also possible that other interactions among effects of other paradigms could occur in ML and not in UL.

Method

Lists.—The transfer effects for four paradigms were studied, each by the use of both UL and ML. If A-B represents the first list, the four paradigms may be symbolized in the second list (UL), or in subgroups of items in the second list (ML), as A-C, A-Br, C-B, and C-D. The last paradigm may be considered the control condition, in that no systematic identities obtain between the items in the two lists (UL) or between the particular subgroup of items for ML.

The materials were paired-associate lists of 12 pairs of two-syllable adjectives. All *Ss* in all conditions with both UL and ML learned exactly the same *second* list. Therefore, any differences in transfer resulting from ML vs. UL cannot be attributed to differences in list or pair difficulty. This second list consisted of the following 12 pairs: *honest—frantic, senior—rotten, certain—aloof, ready—severe, human—unshut, complex—valiant, single—noonday, zigzag—absurd, lukewarm—stubborn, profane—bitter, famous—exact, filthy—rising.* These 12 pairs were selected by use of a table of random numbers from a pool of 24 pairs used by Young (1955) in his low-similarity conditions. The items forming the remaining 12 pairs were used as necessary to construct the first list: for UL, all 12 pairs were used for the A-B, C-D paradigm, none was used for A-B, A-Br, and 12 items were necessary for A-B, A-C, and A-B, C-B.

For ML, with 4 conditions and 12 pairs in a list, each condition was represented by 3 pairs. The use of four groups of *Ss* for ML made it possible to use all 12 pairs in the second list for each paradigm. Thus, in UL, four groups of *Ss* were needed, each group having a given paradigm, and this paradigm held between all 12 pairs in the two lists. In ML, four groups were also used, and for all four groups combined, the relationship between the first and second lists was exactly the same as for UL.

Procedure.—The first list was presented until one perfect trial was attained. After 1 min. the second list was presented for 10 trials. Learning was by the anticipation method, with the stimulus being presented for 2 sec. and the stimulus and response together for 2 sec., as timed by a Hull-type memory drum. The intertrial interval was 4 sec. Three different orders of the lists were used to minimize serial learning. In ML the three pairs representing a given paradigm were not grouped; that is, the ordering of the pairs was such that those representing one paradigm were mixed up with those representing other paradigms. In neither ML nor UL was *S* instructed concerning the nature of the relationships between the two lists.

Four groups of 18 *Ss* each learned under the ML procedure and four other groups of the same number learned under the UL procedure. An order of conditions was made up such that each of the eight conditions was represented once by each block of eight *Ss*. Within each block the ordering was random. The *Ss* were simply assigned in order of their appearance at the laboratory. All *Ss* had previous experience in verbal-learning experiments.

It should be emphasized that all *Ss* learned the same second list. Therefore, it is of little or no consequence whether or not first lists differed in difficulty; this would be true for comparisons within UL or within ML, and it would be true for comparisons between UL and ML. In actual fact, for eight groups, the over-all mean number of trials to learn the first list was 12.7 trials, with a range of from 10.5 to 15.2. An analysis of variance yielded an F of 1.46, which is far from significant. This does not mean, of course, that these first lists did not differ in difficulty since possible ability differences among groups and possible list differences are confounded.

Results

The results will be presented first in graphical form and then attention

will be given to statistical analyses. In Fig. 1, the mean total correct responses on the 10 transfer trials are shown for each condition for both UL and ML. For UL, each condition represents the mean performance of 18 *S*s on 12 pairs. For ML, each condition represents the performance of 72 *S*s, each learning three pairs. To coordinate the values for UL and ML, the total correct responses were summed for successive groups of four *S*s each on ML, such that all 12 items representing a condition for UL were likewise represented for each subgroup of four *S*s in ML. Thus, in a sense, 18 *S*s were "constructed" out of the 72 *S*s learning under ML, and this was done for each transfer condition.

In Fig. 1, it may first be noted that differences between UL and ML are very small. The second fact shown in Fig. 1 is that if C-D is considered the control condition, all three other conditions show negative transfer, the amount of this negative transfer increasing from C-B, to A-C, to A-Br.

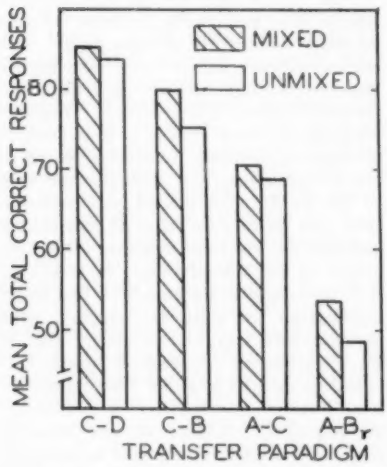

FIG. 1. Total correct responses over 10 transfer trials as a function of mixed and unmixed lists and four transfer paradigms.

In Fig. 1, the values represent the total correct responses over all 10 transfer trials. It is conceivable that a different ordering of the conditions could occur at various stages in learning the second list. That this is not true is shown by Fig. 2, where it is seen that the ordering of conditions remains relatively constant throughout all 10 trials. Also, Fig. 2 shows quite clearly that the differences between UL and ML are of small consequence. The coordination between UL and ML was handled the same way for Fig. 2 as for Fig. 1.

The first statistical analysis consists of a direct test of the results for UL vs. ML. The measure used was the number of correct responses over the 10 transfer trials. These comparisons were made for the total correct for each group of three items. For example, in ML one group of 18 *S*s had a particular group of three pairs for A-Br. These same three pairs were also learned by 18 *S*s in UL. Four t tests could be made for each condition, therefore, and these four tests would exhaust all 12 items used for that condition. With four conditions (A-Br, A-C, C-B, and C-D) and four tests for each condition, a total of 16 t tests was involved. The results were that not a single t was significant, using the .05 significance level. Thus, it may be concluded that for the present materials and conditions, whether UL or ML is used is of no consequence in determining the amount of transfer.

The statistical analysis of results for the UL procedure alone was straightforward. Simple analysis of variance showed conditions to be a highly significant source of variance. Several t tests (using the pooled estimate for the error term) show the following facts: (*a*) If C-D is considered the control condition, A-C and A-Br resulted in significant negative transfer whereas C-B did not. The t values were 6.29, 2.68, and 1.52, respectively, with 34 *df*. (*b*) The amount of negative transfer produced by A-Br is greater than that produced by A-C. The t is 3.61. A comparable analysis performed on the mean correct responses on Trials 1 and 2 did not change the conclusions reached for the analysis of the total correct over all 10 trials.

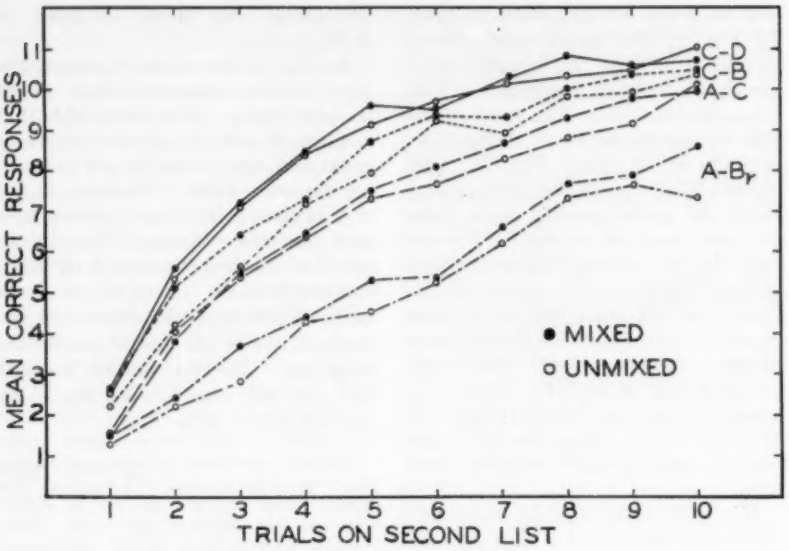

FIG. 2. Acquisition curves over 10 transfer trials as a function of mixed and unmixed lists and four transfer paradigms.

For the statistical analysis of the results for ML alone, direct-difference t tests were used among conditions, each based on 72 *Ss*, with each *S* having three items learned under each condition. Comparing C-D with each of the other three conditions gave t values of 10.09, 4.59, and 2.40, for A-Br, A-C, and C-B, respectively. The only conclusion which differs from those reached for UL is that in ML the C-B condition resulted in significant negative transfer (the t of 2.4 is at the .02 significance level).

For the A-C and A-Br conditions, a count was made of intrusions of first-list responses during the learning of the second. These were counted only if they occurred to the stimulus with which they had been paired in the first list. For A-Br, the total intrusions over 10 trials was 182 for ML and 154 for UL. For A-C, the values were 5 for ML and 4 for UL. Thus, the nature of the list does not appear to seriously affect intrusion-making tendencies.

DISCUSSION

The results have shown that UL and ML produce essentially the same transfer results. Indeed, it is not likely that correspondence would have been higher had the two experiments used the same method (either UL or ML). One implication of these results is that contradictions in the facts of transfer, noted in the introduction, are probably not due to ML vs. UL. A second implication is that the possibility that *S* maintains independent response systems (also suggested in the introduction) in the A-B, A-C paradigm is untenable. The third implication is that it appears that investigators may choose either ML or UL procedures, depending upon circumstances which make one more desirable than another for a particular investigation, without serious concern that the transfer results will be affected by this choice. It is possible, of course, that the present near-identity of results for UL and ML will not hold with all

materials, e.g., consonant syllables, although there is no obvious reason why they should not.

Concerning the transfer effects, the present experiments confirm Porter and Duncan (1953) and Besch and Reynolds (1958) in showing that A-Br produces more negative transfer than A-C. That Mandler and Heinemann (1956) actually found positive transfer with A-Br and negative transfer with A-C may be attributed (as suggested by Besch and Reynolds) to the low meaningfulness of the responses used by Mandler and Heinemann. However, this matter will not be pursued here. It is sufficient to say that it does not appear that the source of contradiction is in the nature of the lists (ML vs. UL).

The one finding that does not fit current theory or previous findings is that the A-B, C-B paradigm produced negative transfer when evaluated against A-B, C-D. It is true that the amount of negative transfer is small, but it is consistent in both experiments and is present throughout most of the 10 transfer trials. Furthermore, there is reason to believe that the negative factor (or factors) may be greater in extent than actually appears in the performance. It is known (Underwood, Runquist, & Schulz, 1959) from other research that teaching *S* responses before learning a list of paired associates will facilitate the learning. In effect, when *S* learns A-B he has learned the responses required in learning C-B; therefore, the learning of C-B should be facilitated. In the present results, this positive factor is more than counteracted by negative effects. The question remains, therefore, as to the source of this interference. The best guess is that it results indirectly from backward learning. It is known (Feldman & Underwood, 1957) that as *S* learns A-B, an association develops between B and A so that B will elicit A with much greater than chance probability. It may be assumed that such backward associations developed in the learning of the first list in the present study; B will have some tendency to elicit A. The *S* then proceeds to learn C-B. As this association develops, it is possible that when C is presented it makes *S* "think of" B, but that B (via backward association) makes him think of A (the stimulus in the first list) leading to an erroneous "conclusion" that B does not go with the present stimulus—C. It would not require many such confusions to produce the negative transfer observed.

Summary

This study was designed to investigate the transfer effects in verbal paired-associate learning using a mixed design (in which subgroups of pairs in the second list form various transfer paradigms with pairs in the first) and an unmixed design (in which all pairs in the second list relate in the same way to the pairs in the first). Four paradigms were studied by each of these two procedures: A-B, A-Br (second-list pairs constitute a repairing of first-list pairs); A-B, A-C; A-B, C-B; and A-B, C-D. The latter paradigm constituted the control condition. For all paradigms for both mixed and unmixed designs all *S*s learned the same second list. The various paradigms were formed by changing the pairs in the first list. Eight groups of 18 *S*s each were used. The first list of paired adjectives was presented until all items were correctly anticipated on a single trial; the second or transfer list was presented for 10 trials.

The results show no difference in the transfer effects as a function of mixed vs. unmixed lists. Thus, the use of various paradigms among subgroups of items within a single list does not produce interactions among the transfer effects for the various paradigms. It was concluded, therefore, that investigators may use either type of design without serious concern that the transfer effects will be different.

Using A-B, C-D as a control, the results show that all paradigms produce negative transfer throughout the 10 trials on the second list. The greatest negative transfer was produced by A-B, A-Br, the next greatest by A-B, A-C, and the least by A-B, C-B. A possible source of interference in the latter paradigm was discussed.

References

Besch, N. F., & Reynolds, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 554–558.

BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, **52**, 360–366.

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

MANDLER, G., & HEINEMANN, S. H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **52**, 39–46.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

YOUNG, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, **50**, 113–119.

(Received August 18, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

INDIVIDUAL PATTERNS OF PHYSIOLOGICAL ACTIVITY AS A FUNCTION OF TASK DIFFERENCES AND DEGREE OF AROUSAL [1]

MORRIS M. SCHNORE [2]

Allan Memorial Institute of Psychiatry and McGill University

A series of studies from this laboratory, using a variety of experimental procedures, have been concerned with the relationship of physiological measures to level of motivation (Bartoshuk, 1955; Malmo & Davis, 1956; Stennett, 1957b; Surwillo, 1956b). As indicated by Malmo, in a recent review (1957), the findings of these studies can be interpreted as being consistent with the formulation of an independent "arousal" (or intensity) dimension of behavior (Duffy, 1951; Hebb, 1955; Lindsley, 1951; Schlosberg, 1954). Furthermore, the data are encouraging with respect to the proposition that physiological measures may serve to quantify this dimension of arousal.

[1] This study was aided in part by research Grant M-1475 from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service, in part by the Research and Development Division, Office of the Surgeon General, Department of the Army, under Contract No. DA-49-007-MD-626; in part by the Defense Research Board of Canada, under Grant No. 9425-04, and in part through a graduate studentship awarded to the author by the National Research Council of Canada. This paper is based on a thesis submitted to the Faculty of Graduate Studies of McGill University in partial fulfilment of the requirements for the Ph.D. degree. The author gratefully acknowledges the generous advice and supervision of R. B. Malmo. He is also indebted to S. M. Feldman who provided helpful criticisms of the manuscript; to J. F. Davis, who designed and supervised the construction of a large part of the equipment; and to W. R. D. Ross and A. K. Bartoshuk, who assisted with the collection of the data.

[2] Now at the University of Western Ontario, London, Ontario.

The main purpose of the present experiment was to attack this general problem from the point of view of individual differences, a topic that has recently been reviewed by Duffy (1957). In part, the aims of the present investigation were similar to those of Lacey and his co-workers in their studies of "autonomic response specificity" (Lacey, Bateman, & Van-Lehn, 1953; Lacey & Lacey, 1958) and those of workers in our own laboratory, in attacking the problem of "symptom specificity" (Malmo, Shagass, & Davis, 1950). The specificity principles hold that individuals differ with respect to which physiological measures show the greatest change under standard conditions of stimulation.

In extending earlier work in this area of individual differences, the present study differs from previous experiments mainly in three respects: (*a*) additional tasks have been employed, (*b*) quantitative variation in arousal (intensity) level was included as a variable in addition to the qualitative differences from task to task, and (*c*) muscle action potentials were added to the autonomic measures. Previous work on specificity with nonpatients had not included measures of activity in the striate musculature.

In addition to the new attack on the problem of physiological specificity itself, there are other purposes of the present investigation that are chiefly bound up with our interest in the concept of arousal. In the first

place, there is a rather important methodological problem to be considered. Some writers have expressed strong doubts about the possibility of finding any systematic relationships between a hypothesized unitary process such as arousal and various physiological measures, because of the lack of consistency in physiological reactions from individual to individual and from one situation to the next. However, the chief hypothesis of the present study is that although individuals will indeed exhibit idiosyncratic patterns of physiological activation, these patterns will be consistent for the individuals in qualitatively different situations, and in those designed to be differentially arousing.

In addition to the matter of individual patterns of physiological activation, the further question of individual differences in what may be called "general arousability" was also investigated. That is, the hypothesis was advanced that whatever their specific patterns might be, some individuals would characteristically show greater activation than others in all of the experimental conditions.

Finally, experimental design included a plan to compare the physiological measures with respect to their apparent power to differentiate meaningfully among the different experimental conditions. This is the question of whether, despite individual differences, some physiological measures appear superior to others in gauging level of arousal.

Method

Experimental Conditions

Individual differences in physiological activity were studied by varying two factors: (*a*) the qualitative nature of the task, and (*b*) the noxiousness of the experimental situation. An attempt was thus made to induce relatively low and relatively high arousal levels while *S*s were doing visual pursuit tracking and while they were doing arithmetic. The conditions were as follows: "low" arousal tracking (LT), "low" arousal arithmetic (LA), "high" arousal arithmetic (HA), and "high" arousal tracking (HT). Every *S* served in all four conditions.

Prior to the main experimental trials, all *S*s had four practice tracking trials. They were instructed to do as well as they could and were informed of their scores after every trial.

In order to cope with the problem of administering a task to *S*s and at the same time achieving a low level of arousal, the often used method of a ruse was adopted. The LT trials were presented to *S* as if they were *not* actually a part of the experiment. After the practice trials were completed, *S* was informed that during the next few trials he would be required to do some arithmetic. Then, before starting the arithmetic, *E* requested *S* to assist him with some adjustments and calibration of the equipment, which purportedly required *S*'s participation.

To increase the level of arousal for the HT trials, *S* was required to track under conditions of intense auditory distraction and was told that he would receive a strong shock if he failed to improve his best score obtained in any of the previous trials. To impress on *S* that this was no hollow threat, a shock was administered 60 sec. before the beginning of the first HT trial. The *S* was informed that if he failed to improve his best score, he would receive a shock twice as strong as the "sample." However, no shock was actually administered after a trial. After the first HT trial *S*s were informed of a score that was about 10% worse than their best score but that they would be given another try.

During the LA trials, *S* was required to perform multiplications of one digit numbers, e.g., $3 \times 8 = ?$ This was, of course, a task involving rote memory.

During the HA trials *S* was required to perform more difficult multiplications, e.g., $17 \times 18 = ?$ In addition to this, he was heckled if he provided a wrong answer or failed to respond within 5 sec., e.g., "This shouldn't take that long!" "Hurry up now!" During the arithmetic trials *S*s were required to hold the tracking pointer at its maximum deflection, in order to equate the muscular demands with those for the tracking.

Subjects

The *S*s were 43 male university students. They were all righthanded, and ranged in age from 16 to 23 yr., with a median age of 19.

Apparatus

Tracking system.—The apparatus used for visual pursuit tracking was a modified version of one previously reported by Surwillo (1956). The modification consisted of: (*a*) major changes in circuitry,[3] (*b*) the two microammeters were mounted one above the other, instead of side by side, which slightly increased the difficulty of the task, and (*c*) ball bearings were incorporated for the controller knob shaft to reduce friction.

The display unit of the tracking system consisted of two pointers (indicators of the two microammeters). The axes of these pointers were 10 cm. apart. The position of the top pointer was varied by a function generator at a constant rate from left to right and back, over a total arc of 60°. A complete cycle of this function, from 0° to 60° and back to 0°, lasted 1 min. The *S*'s task was to control the bottom pointer, so that its reading was always equal to that of the top pointer. The torque required for a maximum deflection of the pointer was 5,000 gm.-cm. and the range of knob rotation was 8°.

The controller knob contained a mechanism which permitted continuous recording of *S*'s grip pressure. Surwillo (1956) has provided a detailed description of its construction.

Shock administration.—A DC shock of 100 v. was administered by a single charged-condenser stimulator. A charge in a two-microfarad condenser was discharged through electrodes, which were placed on the calf of *S*'s left leg. The resistance of *S* through the electrodes was not more than 10,000 ohms.

Auditory distraction.—An intense auditory distraction was presented to *S*s as follows. By means of an audio oscillator, tones continuously varying in frequency between 2,000 and 3,500 cycles were recorded on a Gray Audograph disc. An amplified playback of this recording was fed through a speaker, which was located on the wall directly behind *S*.

Physiological Variables

Muscle activity.—Electromyograms (EMG) from the ventral surfaces of right and left forearms (*pronator teres*), and from the right side of the back of the neck (*semispinalis capitis* and *splenius capitis*) were recorded by an Offner Type D Electroencephalograph. Details of the placement and recording technique used in this laboratory have been reported by Davis (1952). The electroencephalograph was also used to record variations in grip pressure. The muscle potentials and the grip pressure variations were integrated once every 4 sec., and recorded as spikes on a separate chart drive.

All EMG measurements were made in terms of microvolts from the integrated records. From preperiods three spikes were measured: 32–28, 20–16, and 8–4 sec. before the beginning of a trial. Ten spikes were measured for every trial: first, 4–8 sec. after the beginning of a trial and every third spike thereafter. For grip pressure the same sampling method was used, except that measurements were taken for trial periods only. Pressure was measured in grams.

Heart rate.—A continuous recording of heart rate was obtained by an electrocardiotachograph designed by Davis.[4] A chest placement was chosen to insure maximum freedom from EMG artifact. This and the subsequent variables were recorded on a third chart drive.

From preperiods three measurements were made, at 30, 15, and 4 sec. before the beginning of a trial. Eight measurements were made during trials, one every 15 sec. Rate was expressed as beats per minute.

Blood pressure.—Systolic blood pressure was recorded using a modification[5] of an apparatus developed by Burns, Elliott, and James (1955). This method of recording required placing only a small cuff on the middle finger of *S*'s left hand. The recording device provided a measure approximately every 4 sec.

Three measurements were made during preperiods, at 30, 15, and 4 sec. before trials. Nine measurements were made during trials; the first, 4 sec., the second, 15 sec. after

[3] The electronic circuitry of this system, which employed AC amplifiers instead of DC, designed by J. F. Davis, is on file in the Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, Montreal.

[4] Circuitry and description of this apparatus are on file in the Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, Montreal.

[5] The modifications consisted of: (*a*) an industrial air-pump was substituted for the aquarium type pump, (*b*) an eclipse Type DO valve was substituted for the handmade solenoid valve, (*c*) in place of a mercury column and platinum wire, a Statham Type P23AA pressure transducer was used with an Edin Type 5110 Strain Gage Amplifier and Calibrator, and (*d*) an automatic recycling feature was provided by the addition of a Mercoid Switch Type DA 31.

beginning of a trial, the others—one every 15 sec. Pressure was expressed in millimeters of mercury.

Respiration.—Respiration was recorded using a standard type tambour and pneumograph made by Phipps and Bird. For the tracking trials rate was determined simply by counting the number of complete inspiration-expiration cycles. For preperiods, 30 sec. before the beginning of a trial were used.

Because during arithmetic trials *Ss* gave the answers orally, rate was determined by sampling those periods when *Ss* did not talk. Five samples, each consisting of three complete inspiration-expiration cycles, were chosen: one right after the beginning of a trial, and the others—one approximately every 30 sec. This method was also used for some tracking trials where regular cyclical activity for the full trial period was lacking. Rate was expressed in terms of the number of inspiration-expiration cycles per minute.

Palmar conductance.—The circuit used for recording DC skin resistance in this laboratory has been previously reported by Stennett (1957a). This technique involves a standard voltage comparator bridge with a constant current ($\pm 2\%$) over the range of resistances observed in this study. Resistance was measured between the palm and dorsum of the left hand. The *S* was grounded only through the lead from the dorsum. Resistance values were converted into micromhos and all computations were done on these conductance values.

One measure was taken during preperiods, 4 sec. before a trial. Five measurements were made during trials, the first, 4 sec., the second, 30 sec. after beginning of a trial, the others—one every 30 sec.

Skin temperature.—Variations in the temperature of the skin were recorded using a Yellow Springs Instrument Co. Model 45A Tele-Thermometer, a DC amplifier, and a recording galvanometer. The placement of the thermocouple was approximately 4 in. above the wrist on the ventral surface of the left forearm.

One measurement was made during preperiods, 15 sec. after the beginning of a trial. Three measurements were made during trials, one at the beginning, one in the middle, and one at the end of a trial. Temperature was recorded in degrees centigrade.

Testing Procedure

Upon arrival at the laboratory each *S* had a 30 min. interview with *E*. Biographical data were taken, the tracking task was explained, and reassurance given concerning physiological recording. At the end of the interview the short form of the Manifest Anxiety Scale (Taylor, 1953) was administered.[6]

After the interview *S* was taken to the testing room and seated in a special chair with adjustable arm rests, which helped to reduce postural variations from *S* to *S*. Electrodes and attachments were applied, after which routine checks were made on the operation of the recording instruments.

Each trial lasted 2 min., with a 3-min. rest period between trials. Instructions to the *S* were presented through a two-way intercom. A warning to assume the tracking position was given 45 sec. before each trial (preperiod). Altogether there were 12 trials: 4 practice tracking trials, and 2 trials each of the four experimental conditions. The entire experimental routine lasted approximately 2 hr. The tasks were always presented in the following order: practice tracking, "low" arousal tracking, "low" arousal arithmetic, "high" arousal arithmetic, and "high" arousal tracking. A balanced order of conditions was not used because that would have introduced complications for any inter-individual comparison of the *Ss*. The length of the experimental session would then vary from *S* to *S*, because after a high arousal condition different *Ss* would require variable time intervals to recover from the effects of such conditions.

Treatment of Data

Means were computed for the preperiod and trial values from the two trials of each condition for each *S*. To make possible a direct comparison among the different variables, which had been measured in different units, the means were converted into *T* scores (Garrett, 1953). For statistical analysis two indicators of *S*'s physiological activity were computed. One represented the *level* of activity of a particular function and consisted of *S*'s *T* score obtained from his trial values. The other indicator represented a *displacement* of a particular function, which was computed by subtracting a preperiod *T* score from the corresponding trial *T* score. However, since it was found that such absolute change values depended upon the preperiod values, a procedure which partialed out the correlation between the preperiod and trial scores was adopted (Lacey, 1956).

[6] This aspect of the study is not reported here.

TABLE 1

COMPARISON OF MEAN PHYSIOLOGICAL ACTIVITY LEVELS AMONG EXPERIMENTAL CONDITIONS

Variable	Condition					
	LT	LA	HA	HT	χ_r^2	P*
Heart rate	80.7	81.7	90.4	90.5	59.6	.001
Blood pressure	114.5	120.7	128.0	131.1	54.0	.001
Palmar cond.	101.5	105.7	107.8	105.7	37.3	.001
Resp. rate	20.8	20.6	22.6	23.2	58.0	.001
Skin temp.	33.52	33.55	33.45	33.34	14.0	.005
Right forearm	38.3	40.4	47.1	49.9	45.2	.001
Left forearm	6.6	6.4	7.4	9.5	23.0	.001
Neck	20.2	20.1	20.1	22.2	11.1	.01
Grip pressure	666	791	775	866	15.4	.005

* One-tailed test.

A level score of a particular function can be said to reflect the combined influence of all antecedent events (including instructions and anticipation) and the actual performance of a task, while a change score reflects the displacement in a particular function associated only with the performance during a task.

RESULTS

Physiological effects of experimental conditions.—Before presenting the results related to the consistency of individual patterns of physiological activity, the over-all physiological effects of the experimental conditions will first be examined. Table 1 contains mean values for the nine measures of level of physiological activity during the four conditions. Freedman's (see Siegel, 1956) two-way analysis of variance for related samples was used to test the reliability of the differences among the conditions. For all variables the differences are statistically significant.

Table 2 presents the mean changes in the different functions associated with each condition. These data parallel those presented in Table 1. Again, it can be seen that larger and statistically reliable displacements are associated with the "higher" than with the "lower" arousal conditions.

Patterning.—The consistency of each individual's pattern of physiological

TABLE 2

COMPARISON OF MEAN DISPLACEMENT IN PHYSIOLOGICAL ACTIVITY ASSOCIATED WITH DIFFERENT EXPERIMENTAL CONDITIONS

Variable	Condition					
	LT	LA	HA	HT	χ_r^2	P*
Heart rate	− 0.2	0.0	+ 8.2	+ 4.6	49.3	.001
Blood pressure	+ 3.1	+ 8.5	+13.1	+ 8.7	33.0	.001
Palmar cond.	− 1.9	+ 2.8	+ 4.0	+ .07	55.2	.001
Resp. rate	+ 2.8	+ 2.8	+ 4.3	+ 4.6	27.2	.001
Skin temp.	0.0	+ 0.01	+ 0.01	− 0.02	—	N.S.
Right forearm	+27.2	+29.4	+38.9	+34.1	43.32	.001
Left forearm	+ 2.4	+ 1.9	+ 3.0	+ 4.3	24.7	.001
Neck	+ 0.4	+ 0.2	− 0.3	+ 0.2	3.2	N.S.

* One-tailed test.

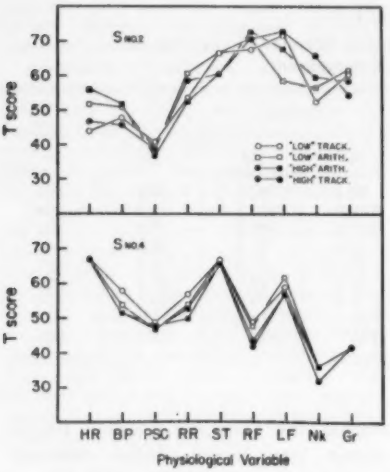

FIG. 1. Cases exemplifying consistency of patterns for physiological activity levels.

activation from condition to condition was evaluated by computing Kendall's (see Siegel, 1956) coefficients of concordance (W). In computing Ws, an individual's pattern was expressed as the rank order of his T scores for the different variables, ranked separately for each condition. For all variables, except skin temperature, high T scores signify high absolute values and large increases. For skin temperature, high T scores were assigned to low absolute values and large decreases because it has been found that anxiety is associated with decreases in finger temperature (Mittelman & Wolff, 1939).

For all *S*s, except one, the Ws are statistically significant ($P < .05$). The Ws range from .995 to .336 with a median W of .801. Two representative cases are presented graphically in Fig. 1. Note the dissimilarity between the patterns exhibited by these two *S*s, while each pattern shows considerable stability from condition to condition.

The Ws, derived from T scores based on the change in activity levels associated with the experimental conditions, are statistically significant ($P < .05$) for 22 *S*s, or 51%. In this case the Ws range from .802 to .227 with a median W of .513. The results based on change in physiological activity are not as decisive as those concerning the level of activity. Fewer *S*s exhibit consistency in their patterns and in general the coefficients of concordance are lower. To investigate the possibility that the low Ws are mainly due to the greater changes in physiological activity associated with the "high" arousal conditions, separate rank order correlations were computed between the two "low" and the two "high" arousal conditions. This was done only for *S*s who did not exhibit significant Ws. Of the 42 such correlations only one was significant at the .05 level. It is thus evident that these *S*s did not exhibit any more consistent physiological patterns during the "low" than during the "high" arousal conditions.

Intra-individual variability of physiological functions.—The cases presented in Fig. 1 also illustrate the fact that there is a considerable amount of intra-individual variability in the different physiological measures. For example, *S* No. 2 is below average in skin conductance, about average in heart rate, but consistently above average in neck muscle action potentials. On the basis of neck muscle action potentials this *S* shows high physiological activation, but in terms of skin conductance his level of activation is low.

If a group of *S*s is exposed to the same experimental conditions, one might expect that all *S*s would not be equally aroused, i.e., some *S*s would show higher over-all physiological activity than others. In terms of

TABLE 3

P VALUES FOR COMPARISONS OF MEAN PHYSIOLOGICAL ACTIVITY LEVELS BETWEEN EXPERIMENTAL CONDITIONS

Variable	Comparison					
	LT vs. HT[a]	LA vs. HA[a]	LT vs. HA[a]	LA vs. HT[a]	LT vs. LA	HT vs. HA
Heart rate	.001	.001	.001	.001	N.S.	N.S.
Blood pressure	.001	.001	.001	.001	.001	N.S.
Palmar cond.	.01	.01	.001	N.S.	.01	.01
Resp. rate	.001	.001	.01	.001	N.S.	N.S.
Skin temp.	N.S.	.01	N.S.	.01	N.S.	N.S.
Right forearm	.001	.001	.001	.001	N.S.	N.S.
Left forearm	.001	.05	N.S.	.001	N.S.	.01
Neck	.01	N.S.	N.S.	.01	N.S.	.05
Grip pressure	.001	N.S.	.025	N.S.	.02	N.S.

[a] One-tailed test.

individual patterns of physiological activity this means that *S*s might differ with respect to the general height of their physiological patterns. That is, even though an *S* is high on one measure and low on another, it is possible that he is high (or low) on most measures, rather than having approximately as many values above the group mean as below. To test this, we would have to see whether the various physiological functions of each *S* were distributed around the group mean, or whether they were instead distributed around a personal mean, which could be either higher or lower than that for the group. A statistical test was done by computing an over-all χ^2. The null hypothesis in this case was that most *S*s have as many values above 50 as below, and therefore, that one cannot differentiate among individuals in terms of their over-all physiological activity levels. Using level scores, the obtained χ^2 was 5.90, which is significant beyond the .001 level. The analogous χ^2 for change scores is 1.60, which is significant at the .05 level. Hence, the null hypothesis must be rejected.

Effectiveness of various measures in differentiating among conditions.—This was evaluated by computing a coefficient of concordance for each *S*. Previously, for the pattern analysis, the *W*s were computed by comparing the rank-orders of physiological measures for the four experimental conditions. In this case, the *W*s were computed by ranking the four conditions for each physiological measure. With respect to the level of physiological activity the *W*s are significant[7] ($P < .05$) for 24 *S*s (56%). The *W*s range from .041 to .733 with a median *W* of .317. For the change in physiological activity the *W*s are significant only for 7 *S*s (16%). The *W*s range from .005 to .581 with a median *W* of .182. However, even the latter distribution of *W*s differs significantly ($\chi^2 = 6.99$, $P < .01$) from the expected distribution if the population coefficient were zero (Lacey & Lacey, 1958). It can be seen, therefore, that the various physiological measures are correlated to a significant degree. Some *S*s, however, show the

[7] There is no reason to expect physiological differentiation between the two "low" and between the two "high" arousal conditions, therefore, the size of these *W*s is generally smaller than if different arousal levels had been induced in each of the four conditions.

predicted changes only in a few functions while others show them in all or most of the functions used in this study.

Finally, the various physiological functions were examined separately to compare their efficiency in differentiating among conditions. Table 3 contains the separate probabilities for mean differences of all measures obtained by making comparisons among the experimental conditions taken two at a time. The first four columns represent the four possible comparisons among the two "low" and the two "high" arousal conditions. In the last two columns are presented comparisons within the "low" and "high" conditions. It can be seen that only four measures consistently differentiate between the two "low" and the two "high" arousal conditions. These measures are heart rate (72% of *Ss* showed faster heart rate during both "high" arousal conditions as compared with the highest of the "low" arousal conditions), systolic blood pressure (72%), respiration rate (64%), and right forearm muscle tension (51%). The relatively small percentage values were obtained because *Ss* often failed to differ in the predicted direction on only one out of the four possible "low"—"high" comparisons. When the above four measures were combined (16 "low"—"high" comparisons for each *S*) one could differentiate reliably ($P < .05$) between the "low" and "high" arousal conditions for 86% of the *Ss*. The other measures fail to differentiate significantly at least between one of the "low" and one of the "high" arousal conditions.

Only three measures differentiated significantly between the two "low" arousal conditions. These were systolic blood pressure, skin conductance, and grip pressure; the values being significantly higher during the arithmetic as compared with the tracking trials. The two "high" arousal conditions also showed significant differences on three measures. However, although left forearm and neck muscle tension were significantly greater during the tracking trials, *Ss* showed significantly higher skin conductance during arithmetic trials.

The probabilities yielded by comparing any two conditions for the amount of change are presented in Table 4. The large number of small probabilities, especially for the measures which showed high differentiation in terms of level scores, indicate

TABLE 4

P VALUES FOR COMPARISONS OF MEAN CHANGES IN PHYSIOLOGICAL ACTIVITY BETWEEN EXPERIMENTAL CONDITIONS

Variable	Comparison					
	LT vs. HT[a]	LA vs. HA[a]	LT vs. HA[a]	LA vs. HT[a]	LT vs. LA	HT vs. HA
Heart rate	.001	.001	.001	.001	N.S.	.05
Blood pressure	.001	.001	.001	N.S.	.001	.001
Palmar cond.	.001	N.S.	.001	N.S.	.001	.001
Resp. rate	.001	.001	.001	.001	N.S.	N.S.
Skin temp.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
Right forearm	.001	.001	.001	.001	N.S.	N.S.
Left forearm	.005	.025	N.S.	.001	N.S.	.005
Neck	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.

[a] One-tailed test.

that the probabilities in Table 3 are a result of the different arousal functions of the two types of conditions rather than only to the differences in the "alerting" response itself (instructions, anticipation).

Discussion

The hypothesis that during qualitatively different stimulus situations *Ss* will exhibit idiosyncratic but highly stereotyped patterns of somatic and autonomic activation is clearly supported by the data. The same *S* might show below-average levels of activity in some functions, average in others, and above-average in still others. But no matter what the shape of the idiosyncratic pattern, *Ss* tend to respond with their specific hierarchy of activation to situations which require different overt responses. Even when variations in experimental situations produce increases in the over-all level of activation, significantly stereotyped physiological responses are observed.

These findings are consistent with those previously reported by Lacey et al. (1953) and Lacey and Lacey (1958) who formulated the principle of "response specificity" using autonomic variables. In view of the present results such a principle would seem to apply not only to the functioning of the autonomic nervous system but to the skeletal-muscle system as well.

The fact that during the high arousal trials a general increase in physiological activity was observed would seem to support the notion of a general arousal state. However, any given level of arousal is reflected differentially in different individuals. Such differentiation is contrary to the traditional broad classifications, such as autonomic-somatic, sympathetic-parasympathetic, since differentiation did not follow such classifications but was evident even within cardiovascular and somatic systems themselves. This lack of homogeneity indicates a need for a new approach in our thinking about the functioning of the nervous system. Recently, Wenger (1957), in discussing some new data representing the activity of the autonomic nervous system, drew a similar conclusion.

In general, the results of this study are in accord with the notion of an arousal continuum and support the suggestion that peripheral physiological variables can be used to measure changes in arousal. Higher levels and greater increase in physiological activity were observed during the "high" than during the "low" arousal conditions. One might, however, pose the question whether varying degrees of arousal were actually induced in this experiment. To avoid circularity, one obviously can not use the physiological data themselves in support of an affirmative answer. A recourse to other observables is necessary.

The arousal theory implies that degree of arousal is a function of task difficulty and intensity of stimuli (physical or symbolic) (Duffy, 1951, p. 32). That is, other things being equal, during difficult tasks arousal is expected to be higher than during easy tasks, and noxious situations are expected to produce higher levels of arousal than relatively innocuous situations. It appears that the experimental conditions satisfy these requirements adequately. Consequently, one can interpret the data as supporting the prediction that increases in arousal will be associated with increased physiological activity.

Before continuing with this question of applying physiological measures to gauge level of arousal, an important methodological distinction deserves consideration: it is essential to be clear whether the comparisons of arousal level are inter- or intra-individual in character.

In comparison with others one *S* may show relatively high heart rate and relatively low muscle tension. The statement that an *S*'s heart rate is higher than the heart rate of others, is, of course, an interindividual comparison. Similarly, the statement concerning his relatively low muscle tension is also based on comparison with other *Ss*. That an *S* may be high relative to others in one physio-

logical function and low (again relative to others) in another physiological function is what we should expect from findings of idiosyncratic patterns of physiological activity. Such individual differences would also account for the low intercorrelations among physiological measures that have been reported by Ax (1953) and others (Lacey & Lacey, 1958; Malmo & Davis, 1956; Malmo & Smith, 1955).

But there is a point that, curiously enough, appears to have been overlooked by some writers. Despite his idiosyncratic pattern, an individual placed in an arousing situation will, according to the results of the present study, probably show a general increase in most physiological functions. The hypothetical case cited above would be expected to show increase in heart rate and also in level of muscle tension, even though, in comparison with others, muscle tension remains low, and heart rate is still high. It is possible, therefore, to obtain a zero correlation between any two variables when the correlation is computed among *Ss*, while the correlation between the same two variables can be perfect if computed within an *S*. Present data indicate that, although the correlations are generally low among *Ss*, there is considerably closer correspondence among different physiological measures within the same *S*.[8] For these reasons, self-control design (in which the same *Ss* are employed in all conditions) appears superior to design employing independent groups.

Another methodological point impressed on us by present findings is that whether comparisons are intra- or inter-individual in character, there are significant tactical advantages in employing several physiological measures rather than in relying on only one or two. The large interindividual differences in activation, seen in all physiological functions, indicate that an individual's level of arousal, in a given situation, can not be reliably inferred from a single physiological measure (somatic or autonomic). One would obtain a different ordering of *Ss* with respect to degree of arousal, depending on which physiological measure was chosen as the criterion. Although any one physiological measure appears to be an unreliable criterion of the general arousal level, by using a number of measures in combination, it seems highly probable that one can differentiate reliably among individuals. Some individuals show characteristically greater physiological activation than others in response to standard situations. In addition, even for intra-individual comparison the use of several physiological measures seems to be indicated, since individuals vary considerably in the amount of correspondence that exists among physiological measures and the significant concordance coefficients seldom approach 1.00.

Recently, several studies have been reported where the relationship between physiological activity on one hand, and MAS scores and performance data on the other hand, has been investigated (Beam, 1955; Berry & Martin, 1957; Calvin, McGuigan, Tyrrell, & Soyars, 1956; Raphelson, 1957). The results of this study suggest that the failure to find significant correlation between physiological activity and MAS, and performance might be expressly due to the undependability of a single physiological measure as an indicator of general physiological activity.

In addition to the individual differences discussed above, it was found in the present study that, for purposes of distinguishing among arousal levels, some physiological functions seem to be more adequate than others. Such differences among the physiological measures exemplify the differential sensitivity of the various functions. It appears that a given change in the general arousal state might be reflected consistently in some peripheral bodily functions and not in others. While heart rate, systolic blood pressure, respiration rate, and right forearm muscle tension differentiated consistently among conditions, skin con-

[8] Such correspondence among physiological measures is more pronounced with respect to the direction of change, rather than the amount of change.

ductance, skin temperature, left forearm muscle tension, and grip pressure failed to differentiate at least between one of the "low" and one "high" arousal condition. By combining the first four measures more reliable differentiation between the arousal levels was possible than on the basis of any one of these measures alone.

In terms of the total arousal continuum, arousal levels exhibited by *Ss* of this study would seem to represent moderate degrees of arousal. Such an inference is based on two considerations. First, even during the "low" arousal conditions *Ss* were required to perform a task. Second, the "high" arousal conditions were not of great enough intensity to produce a breakdown of performance; in fact, there were no striking changes in overt behavior. Accepting heart rate, blood pressure, respiration rate, and right forearm (active) muscle tension as criteria, it was possible to discriminate reliably between the "low" and "high" arousal conditions. This study indicates, therefore, that changes in the degree of arousal in the middle range of the total arousal continuum can be differentiated effectively by the combined use of these four physiological functions. It is very likely that different criteria might have to be used, depending on the amount of change of arousal one wants to detect, and in which part of the arousal continuum the given changes occur.

Summary

The chief purpose of the present study was to investigate the consistency of individual differences in physiological activity. In addition, some of the problems related to the measurement of arousal were also evaluated. Nine physiological functions were recorded. Variations in experimental situation were produced by employing tracking and arithmetic tasks with two different levels of arousal for each task.

Results showed that during qualitatively and quantitatively different stimulus situations individuals exhibited idiosyncratic but highly stereotyped patterns of somatic and autonomic activation.

The present results further support earlier work which found physiological measures singularly useful in providing objective and reliable indicants of arousal. Fifty-six per cent of *Ss* showed significant correlation among the physiological measures. There were four measures (heart rate, blood pressure, respiration rate, and right forearm muscle tension) which consistently differentiated between high and low arousal conditions. Present evidence suggests, therefore, that these four measures may be maximally useful as indicants of arousal, at least within the range of arousal levels studied. Although data from other experimental situations are required in order to reach a firmer conclusion concerning the relative usefulness of these measures, it seems certain that several indices are superior to only one or two in gauging level of arousal.

References

Ax, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, **15**, 433–442.

Bartoshuk, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.

Beam, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–551.

Berry, J. L., & Martin, B. GSR reactivity as a function of anxiety, instructions, and sex. *J. abnorm. soc. Psychol.*, 1957, **54**, 9–12.

Burns, B. D., Elliott, H., & James, C. E. Continuous recording of respiration, heart rate, and systolic blood pressure. *Canad. Serv. med. J.*, 1955, **9**, 509–514.

Calvin, A. D., McGuigan, F. J., Tyrrell, S., & Soyars, M. Manifest anxiety and the palmar perspiration index. *J. consult. Psychol.*, 1956, **20**, 356.

Davis, J. F. *A manual of surface electromyography.* Montreal: Lab. for Psych. Stud., Allan Mem. Inst. of Psychiat., McGill Univer., 1952.

Duffy, E. The concept of energy mobilization. *Psychol. Rev.*, 1951, **58**, 30–40.

Duffy, E. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, **64**, 265–275.

Garrett, H. E. *Statistics in psychology and education.* (4th ed.) New York: Longmans, Green, 1953.

Hebb, D. O. Drives and the C. N. S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–253.

LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123–164.

LACEY, J. I., BATEMAN, D. E., & VAN LEHN, R. Autonomic response specificity. *Psychosom. Med.*, 1953, **15**, 8–21.

LACEY, J. I., & LACEY, B. C. Verification and extension of the principle of autonomic response-stereotypy. *Amer. J. Psychol.*, 1958, **71**, 50–73.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MALMO, R. B., & DAVIS, J. F. Psychological gradients as indicants of "arousal" in mirror tracing. *Canad. J. Psychol.*, 1956, **10**, 231–238.

MALMO, R. B., SHAGASS, C., & DAVIS, F. H. Specificity of bodily reactions under stress. *Proc. Ass. Res. nerv. ment. Dis.*, 1950, **29**, 231–261.

MALMO, R. B., & SMITH, A. A. Forehead tension and motor irregularities in psychoneurotic patients under stress. *J. Personal.*, 1955, **23**, 391–406.

MITTELMAN, B., & WOLFF, H. G. Affective states and skin temperature: Experimental study of subjects with "cold hands" and Raynaud's syndrome. *Psychosom. Med.*, 1939, **1**, 271.

RAPHELSON, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18.

SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

STENNETT, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG Clin. Neurophysiol.*, 1957, **9**, 131–138. (a)

STENNETT, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61. (b)

SURWILLO, W. W. A device for recording variations in pressure of grip during tracking. *Amer. J. Psychol.*, 1956, **68**, 669–670. (a)

SURWILLO, W. W. Psychological factors in muscle-action potentials: EMG gradients. *J. exp. Psychol.*, 1956, **52**, 263–272. (b)

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WENGER, M. A. Pattern analysis of autonomic variables during rest. *Psychosom. Med.*, 1957, **14**, 240–244.

(Received August 15, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

CHANGES IN GSR TO A SINGLE STIMULUS AS A RESULT OF TRAINING ON A COMPOUND STIMULUS[1]

WILLIAM W. GRINGS AND VSEVOLOD N. SHMELEV

University of Southern California

The study reported here is one of a series exploring relations among learned reactions to simple stimulus compounds and their elemental component stimuli. Classical conditioning of the GSR has been used throughout, with colored dots (single or paired) as visual conditioned stimuli and electric shock as the UCS. The major relations previously studied included response transfer from single stimuli to pairs (Grings & O'Donnell, 1956), transfer from paired stimuli to components, and the influence of preconditioning variables (Marks & Grings, 1958).

The present experiment evaluates the assumption that training of stimulus components as members of a compound may alter the response magnitudes previously conditioned to the components separately. The specific hypothesis tested is that when an organism is trained to respond to individual component stimuli, and is subsequently given training on a compound composed of these components, the response evocation potentials of the components will undergo a change.

It is predicted further that when the compound stimulus is reinforced, the effect on component response tendencies will be opposite in direction to the effect when the compound stimulus is presented without reinforcement. On the basis of simple transfer from components to compounds it would be predicted that reinforcement of a compound would add to the response potentials of the components contained in the compound. On the other hand, presentational experience (nonreinforcement) of the compound would lead to extinction of the component reactions.[2]

Briefly, the experimental procedure was as follows. GSR were established to four visual stimuli by classical conditioning techniques. Two of the stimuli were then combined in a compound on which further training was given. The other two were used as control stimuli. For one random group of *S*s the training to the compound was with reinforcement present (hereafter called *reinforcement experience*); whereas with a second group of *S*s training to the compound was with reinforcement absent (hereafter called *presentational experience*). Finally a test was administered on all four individual stimuli to note the effects of compound training.

To the knowledge of the writers the only experimental evidence on this question comes from the laboratory of Bekhterev. As reported by Razran (1939) this group found that the conditioning of a compound was achieved at the expense of the gradual elimination of component conditioning, the weaker components being

[1] This study was supported by a grant (G-925) from the National Science Foundation.

[2] An alternative hypothesis to those posed here might be that the *S*s would readily discriminate between presentation of components and compounds, and therefore the experience with compounds would have no effect upon the response potential of the individual components.

eliminated first, the stronger last. Light and sound stimuli were employed with human *Ss* and withdrawal to electric shock was the response.

METHOD

Subjects.—Forty volunteers, all undergraduate students of the University of Southern California, were employed as *Ss*. Each was reimbursed for his participation. All had normal color vision as tested by the Ishihara test.

Apparatus.—Conditioned stimuli were small colored dots produced by projection through Wratten filters onto a ground glass screen in front of *S*. The four colors, red, green, blue, and yellow, were obtained from standard Wratten filters and were matched for brightness. Time relations among the stimuli were controlled electronically with the CS preceding the UCS by .5 sec. and remaining on .5 sec. after the onset of the UCS. The UCS was a condenser-discharge electric shock to the forearm. The GSR was picked up as a DC resistance change from the palm and the back of the nonpreferred hand. A more detailed account of the basic instrumentation and laboratory situation is given in a previous study (Grings & O'Donnell, 1956).

Preliminary procedure.—The *S* was seated in front of a screen built into a partition that separated *E* and the equipment from him. After a preliminary interval of 15 min. for electrode stabilization and apparatus calibration each of the four colored stimuli was presented to *S* in a random order to acquaint him with the stimuli. The shock was then adjusted in intensity to achieve a double criterion: the consistent production of a GSR, and a verbal expression of slight discomfort by *S*. After three shock sensitization trials, *S* was instructed in his "task" for the remainder of the experiment. This was to depress an ergograph trigger with the index finger of his preferred hand when *E* said the word "ready," and to keep a steady pressure on the trigger until the visual stimulus went out. The *S* had previously been informed that the purpose of the experiment was "to determine the nature of your skin electricity changes while we present colored stimuli on the screen in front of you." He had also been told that the shock provided a "standard sized change" for use as a reference value.

Main procedure.—The major experimental sequence contained three parts: (1) a period of training on single stimuli, (2) a period of training on a compound or pair of stimuli, and (3) a test series on single stimuli. For Parts 1 and 3 all *Ss* were treated alike, whereas for Part 2, *Ss* were assigned randomly into experimental subgroups. The following discussion of procedure details is organized according to the chronology of the three parts listed above.

In Part 1, the period of training on single stimuli consisted of six presentations of each of four colored lights, three colors being reinforced (paired with shock) and one not reinforced. Presentation order was controlled by assigning each color randomly to two positions in each block of eight trials. The assignment of colors to reinforcement-nonreinforcement conditions was varied systematically from *S* to *S*. Except for the first 11 *Ss* the component training segment was concluded by a single presentation of each color without reinforcement.

Part 2 was devoted to training on a stimulus compound composed of two of the previously reinforced single colored dots presented side by side as a pair. One member of the pair was chosen at the outset to represent the effects of compound training for purposes of data analysis. That component was referred to as the *included* stimulus. The third reinforced component from Part 1 was not presented during Part 2. It was used later as a reference stimulus and was referred to as the *nonincluded* stimulus. The main dependent variable in the study was the difference between responses to the included and nonincluded stimuli during the final test segment. The fourth color, which had been presented without reinforcement during the component training segment, was also not presented in Part 2. It was used later only as a control stimulus to evaluate the presupposition that learning had occurred.

For the assignment of compound training conditions to the *Ss* the total sample was divided randomly into two groups. Group I (reinforcement group) received eight reinforced trials on the compound stimulus while Group II (presentational experience group) received eight nonreinforced trials. In order to provide data from which to plot curves of GSR to the included stimulus during the course of compound stimulus training, each major group was divided randomly into two subgroups of 10 *Ss* each. Subgroup A in each case received compound training without interruption, while Subgroup B had a test trial on the included stimulus interspersed between the fourth and fifth compound training trials.

Part 3, the final test series, included the presentation of each component stimulus

twice without reinforcement. This part immediately followed the compound training part and for purposes of analysis was broken up into a first posttest which consisted of the first presentation of all four colors in a random order, and a second posttest which consisted of the second presentation of the same stimuli in a second random order.

At the end of the test series *S* was asked to report his reactions concerning the nature of the experiment. These interviews were either recorded on tape or summarized and written down by *E*.

Results

The first step in the treatment of the data involved the transformation of the GSR into the unit square root of conductance change. This unit had been found in previous studies (Grings & O'Donnell, 1956) to have certain advantages. The major analysis to be presented here concerns a dependent variable which is the difference in response to the two crucial stimuli, the included stimulus and the nonincluded stimulus, as previously defined. Both of these stimuli were originally reinforced and one of them, the included one, was given compound training while the other was not. The first evaluation involved a *t* test of the difference in the dependent variable due to the kind of compound training (reinforcement vs. presentational experience). These results are presented in Table 1 where it can be seen that, for the group with the reinforced compound (Group I), on the first test trial the response to the included stimulus is larger than that to the nonincluded stimulus (the difference variable is positive). With the presentation group (Group II) the reverse is true—the response to the included stimulus is smaller than the response to the reference stimulus (the difference variable is negative).

TABLE 1

Mean Differences in GSR to Included and Nonincluded Stimuli for the Two Compound Training Groups[a] (N = 20 for each group.)

Group	First Posttest	Second Posttest
I. Reinforcement of compounds	+ 5.95	−14.50
II. Presentation of compounds	−14.75	− 3.15
t ratio	2.42	1.55
P	.02	.15

[a] Unit: square root of conductance change in .01 micromhos. In each case the GSR to the reference (nonincluded) stimulus has been subtracted from the GSR to the experimental (included) stimulus.

The group differences on the first test trial in Table 1 are in the direction that would be expected if reinforcement experience with the compound increased the response tendency while presentation experience reduced, or extinguished, such response tendency. It will be noted from the table that the difference between the two groups is significant at the 2% level (two-tailed test). However, it will also be noted that this group difference reverses itself on the second test trial where, for both groups, the response to the included stimulus is less than that to the reference stimulus. This difference is not significant.[3]

The nature of these effects may be seen graphically by observing the magnitude of the experimental (included) stimulus throughout the compound training and posttest periods. Results from the subgroups for which a test trial was introduced midway during compound training are shown

[3] Because the distribution of GSR was moderately skewed even after transformation into square root of conductance change, all *t* tests were paralleled by nonparametric analyses of the same data. For within-*S* comparisons the Wilcoxon matched-pairs signed-ranks test was used and for between-groups the Mann-Whitney *U* test was employed.

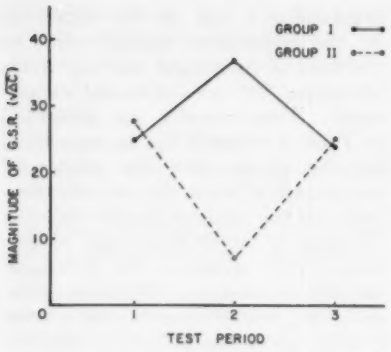

FIG. 1. Comparison of median GSR magnitudes for the *included* stimulus at points: (1) after four trials on compounds; (2) after eight trials on compounds, or the first posttest trial; and (3) on the second posttest trial. $N = 10$, in each group.

in Fig. 1. The first point indicates the response at the middle of compound training; the second point at the end of compound training (or the first test series); and the third point at the second test series. Median values are plotted and support the conclusion that a difference between groups appears at the end of compound training but disappears after one test series. In agreement with Table 1 the difference between the means on Point 2 of the figure is significant at the 5% level ($t = 2.20$; $df = 18$).

Interpretation of the above tests rests on two basic presuppositions: that Groups I and II were not different at the outset of compound training and that learning of some sort had occurred to the reinforced components during the pretraining period on single stimuli. To test the first of these, the differences between the GSR on the included and reference stimuli were computed for the test trial following the initial component conditioning. When this value (which is comparable to the variable in Table 1) was compared for the two groups, no difference was found ($t = .63$).

Similarly, to evaluate the presupposition of discrimination learning, a comparison was made of the reinforced reference stimulus and the fourth stimulus component during the two posttest series when both were presented. If effects of discrimination learning persisted until the posttest period, the GSR to the reinforced reference stimulus (the so-called nonincluded stimulus) should be greater than the GSR to the fourth component which had been presented previously only during training on individual stimuli and had not been associated with shock. Groups I and II were combined for this comparison and a significant difference in the expected direction (supporting the assumption of discrimination learning) was found ($t = 2.01$; $df = 38$).

DISCUSSION

The analysis of results presented above suggests the conclusion that when stimuli are originally trained as individual components and subsequently trained as compounds, their response potential (as components) may undergo a change as a result of the compound training. At this point it appears best to conclude that positive transfer results from reinforcement training and negative transfer (or extinction) results from presentational or nonreinforcement experience. It is also apparent that these effects, at least for the GSR, are not stable and are influenced readily by extinction factors.

Transfer of the type described is quite what one would expect from principles of classical conditioning. The suggestion of reversal from first to second test trials, however, could appear to be at odds with such an explanation. Emphasizing the fact that human *S*s are involved, the writers are inclined to rely on factors of voluntary response, set

change, or change in perception of the experimental situation. With self-instructing, verbalizing organisms it is presumed that during preliminary training *Ss* were led to anticipate that certain colors would be followed by an electric shock. During reinforcement training on compounds the anticipation of shock was being strengthened for the included stimuli, whereas during nonreinforcement training on compounds the anticipation of shock was being weakened. As a result, on the first test trial after compound training the effects of simple reinforcement-nonreinforcement transfer were evident. However, the experience of all stimuli without shock on the first test trial created in *S* an ambiguous perceptual situation for subsequent test trials.

The empirical fact of rather dramatic change in GSR magnitude from first to second extinction trials has been reported before (e.g., Grant & Schiller, 1953). The importance of verbal and self-instruction factors on the learning and extinction of this particular response has also long been emphasized (Cook & Harris, 1937; Humphreys, 1940; Woodworth, 1947).

If the above "perceptual" conditions are true, other evidence might be expected from the data. One such indication was explored but led to inconclusive results. It employed information from the postexperimental interview on the basis of which *Ss* could be classified according to whether they were *aware* of stimulus relations at the conclusion of the experiment or were *unaware* of them. Assuming that unaware *Ss* were reacting less to their own verbal conception of what was going on than were the aware *Ss*, it was predicted that the unaware *Ss* would show less reversal between the first and second test series and that such an effect should be present only in Group I (since only that group experienced a change in reinforcement experience at the time of final test). A tabulation of data according to these awareness conditions showed that the average change in the dependent variable from first to second test series was largest for the aware *Ss* of Group I and smallest for the unaware *Ss* of the same group. A Mann-Whitney test of this difference, however, was not significant

Summary

A test was made of the prediction that when an organism is trained to individual component stimuli and is subsequently given training on a compound composed of these components, the response evocation potentials of the components may undergo a change. The magnitude of this change was studied as a result of (*a*) reinforcement of the compound and (*b*) mere presentation of the compound.

The results, with 40 *Ss* and GSR conditioning to visual stimuli, support the expectation of a change in the direction of greater magnitude of response with reinforcement experience in the compound and less magnitude of response with presentational experience. The fact that these results disappeared between first and second test series was discussed as due to a combination of reinforcement and perceptual factors.

References

Cook, S. W., & Harris, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

Grant, D. A., & Schiller, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.

Grings, W. W., & O'Donnell, D. E. Magnitude of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1956, **52**, 354–359.

Humphreys, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.

Marks, A., & Grings, W. W. Preconditioning factors in compound stimulus conditioning. Paper read at Western Psychol. Ass., Monterey, Calif., April, 1958.

Razran, G. H. S. Studies in configural conditioning: I. Historical and preliminary experimentation. *J. gen. Psychol.*, 1939, **21**, 307–330.

Woodworth, R. S. Reinforcement of perception. *Amer. J. Psychol.*, 1947, **60**, 119–124.

(Received August 15, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

EFFECTS OF THREAT OF SHOCK, DISTRACTION, AND TASK DESIGN ON PERFORMANCE [1]

ROBERT E. MURPHY [2]

Johns Hopkins University

This research is an attempt to integrate two different but closely related lines of research in psychology. The first covers the investigations of the effects of psychological stress on performance; the second covers the studies of the effects of different task designs on operator efficiency.

Stress studies and task design studies have not usually been thought of within the same framework. Investigations of the effects of stress on performance have typically emphasized the manipulation of factors external to the task to be performed. Subjects have been threatened with failure, given false norms, distracted by all sorts of extraneous stimuli, and have been placed in various situations which have seemed to experimenters to be stressful. Though there are exceptions (Gates & Rissland, 1923; Hurlock, 1924; Verville, 1946), the great majority of the studies reviewed by Lazarus, Deese, and Osler (1952) report that performance is impaired as a result of psychological stress.

In contrast to the stress studies, research on the effects of task designs on operator efficiency has been mainly concerned with the manipulation of the task itself. The critical consideration in these studies is the nature of the demands made upon the operator by the task design. For example, in the design of visual displays, some of the important considerations seem to be the amount of search required to perceive relevant cues, the amount of interpolation required, the degree to which elements in the display conform to expected or learned relationships, and the degree to which cues are confused. The efficiency of the operator varies as the number and complexity of the demands made upon him are changed. Lazarus, Deese, and Osler (1952) have suggested that a task itself may be inherently stressful due to such demands. This type of stress factor is not unlike that which Conrad (1950) has referred to as "load stress."

In line with this analysis, then, two types of stress may be distinguished on the basis of the conditions manipulated to induce them—that induced by varying conditions external to the task, and that induced by the task design itself. This distinction raises the question as to what might happen when these two types of stress are varied concomitantly. On the basis of previous findings, stress induced by factors external to the task should result in a performance decrement. One also expects differences in task design to produce differences in performance.

The issue of primary concern in the present study, however, is the nature of the interaction of these two sources of stress. It seems possible that combining two forms of stress—external and task induced—will produce a combined effect which is not simply additive but appears in the

[1] This article is based on a dissertation submitted to Johns Hopkins University in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to W. R. Garner and L. S. Otis for their many helpful criticisms and advice.

[2] Now at the Wright Air Development Center, Wright-Patterson A. F. B., Ohio.

form of an interaction, such that the differential effect of one type of stress will be increased with the addition of the other type. This suggestion is supported by studies of Chapanis (1954), Conrad (1950), and Mackworth (1952) which show that the effect of a given increment of stress is greater if it is added to a higher rather than a lower existing level of stress.

The two most commonly used forms of external stress are those which follow the "anxiety" paradigm and those which follow the "distraction" paradigm. The anxiety types are those in which the individual anticipates physical or psychological pain. The distraction types are those in which the individual is deliberately bothered, disturbed, or distracted during the performance of his task. Accordingly, an experiment was designed to test the effects of distraction and threat of electric shock on a task performance. Four different task designs were used. The experiment was designed so that the effects of distraction and the effects of threat of shock could be tested alone and in combination for each of the task designs. The hypotheses specific to this experiment were as follows: (*a*) The differences in performance obtained with different task designs are greater under threat of shock than under nonthreatening conditions. (*b*) The differences in performance obtained with different task designs are greater under distraction than under nondistracting conditions.

Method

General Procedure

Subjects.—The *S*s were 80 male undergraduate students from an introductory psychology class at Johns Hopkins University.

Task materials.—The materials used in this experiment were four decks of cards. Each deck was made up of 32 square cards, 5 in. on a side. The cards were photostatic duplications of range-ring patterns, covered with a transparent plastic which made the cards rigid and easy to handle. Each deck contained a different pattern of eight rings, and on each card one of the rings was marked with a large **X**. The marks were arranged so that each ring was marked four times in each deck, once in each quadrant. The range-rings had an outside diameter of 4.0 in. and were separated from each other by .25 in. The rings were solid or dashed black lines against a white background. The four different range-ring patterns are shown in Fig. 1.

Preliminary trials.—The *S* was seated at a table opposite *E* and was handed a preliminary deck of 32 cards. The cards in this preliminary deck were identical in every respect with the task cards, except that instead of a pattern of range-rings, a number from 1 to 8 was marked in the center of the card. The *S* was instructed to deal the cards from the top of the deck as rapidly as possible and to call off the numbers marked in the center of each card. At the signal, "ready—go," *S* ran through the preliminary deck. The *S* was given two trials with the preliminary deck, and the time required to deal the deck on the second trial was recorded by *E*. The time score for the second preliminary trial will be referred to throughout this paper as *S*'s preliminary score.

Immediately after the preliminary score

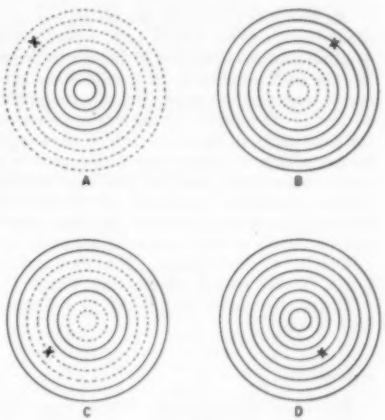

FIG. 1. Range-ring patterns used in the experiment.

was recorded, *S* was handed one of the decks of task cards. He was instructed to deal the cards off the top of the deck as rapidly as possible and to call off the numbers of the marked range-rings. He was further instructed to avoid making errors. Each time *S* made an error, he was stopped and his attention was called to the error. He was then permitted to proceed with the rest of the cards in the deck. As soon as one deck was completed, *S* was given the next deck (with a different pattern of rings) and so on, until all four task decks were run through once in this manner.

Each *S* next ran through all four decks a second time, after being advised that he would no longer be told if he made an error. For this second series, the cards in each deck were arranged in reverse order. The *E* recorded the time required to deal each deck and the number of errors made.

After completing the second preliminary series, each *S* was told that he would be required to deal all four decks two more times. He was asked to try to do better than the average score of the *S*s who had dealt the same cards during the previous week. He was further told that he would be paid 5 cents for each second that his average time score for the next two series bettered the average time score for the previous week, but that 5 cents would be deducted from his winnings for each of the average number of errors he made.

Special Procedure

At this point, each of the 80 *S*s was randomly assigned to one of four experimental conditions so as to provide 20 *S*s per condition. Within each condition, *S*s were further randomly assigned to one of four subgroups of 5 *S*s each. The order of the presentation of the four decks of cards to each of these subgroups within an experimental condition was counterbalanced. Each of the four experimental conditions consisted of a different combination of two levels of threat of shock and two levels of distraction, as follows:

Cond. I—No threat, No distraction; Cond. II—No threat, Distraction; Cond. III—Threat, No distraction; Cond. IV—Threat, Distraction.

Variables

Threat of electric shock.—There were two levels of this variable: (*a*) no threat, and (*b*) threat of an electric shock for every error made during the final two series. To produce the threat of shock, an electrode was strapped about 3 in. above the left ankle of *S* immediately after he had completed his last preliminary trial. He was told that he would receive an electric shock for every error he made during the final two series of task cards. Actually, *S* received a mildly painful shock for each error during only the first of the final two series. The *S*s were not shocked for errors made during the second series.

The shock apparatus was a variable coil inductorium operated by a 4.5-v. DC source. The intensity of the shock was set at the average level judged to be "just painful" by a group of 16 *S*s in a preliminary experiment. The duration of the shock was .5 sec. The same setting was used for all *S*s.

Shock for an error during the first of the final two series had the added effect of providing information about errors to those *S*s in the threat condition. As a control for the influence of this effect, *S*s in the nonthreat condition were provided information about errors made during the first of the final two series by a .95-w. red light which was placed on the table in front of *S* and which flashed for .5 sec. every time an error was made.

Distraction.—There were two levels of distraction: (*a*) no distraction, and (*b*) distraction produced by presenting aurally to *S* a series of random numbers from 1 to 8. The random numbers were recorded on a tape and presented to *S* through a set of earphones in a moderately loud conversational tone (approximately 75–80 db above threshold). After *S* had completed his last preliminary trial, the earphones were adjusted on him, and he was given the final two series. The recorder was turned on about 3 sec. before *S* was given the signal to begin dealing a deck of cards, and it was turned off shortly after each deck had been completed.

Task design.—There were four task designs. These are represented by the four range-ring patterns shown in Fig. 1. Garner, Saltzman, and Saltzman (1949) have previously shown that these four patterns produce significant differences in performance.

Criteria of performance.—Two criteria were used: (*a*) time to deal a deck of 32 cards, and (*b*) number of errors made.

Summary of Experimental Design

The design used in this experiment is essentially Lindquist's (1953) Type III design. Eighty *S*s dealt all four decks of cards. There were four combinations of threat and distraction conditions with 20 *S*s in each combination. The design also provided for

TABLE 1

PRELIMINARY SCORES FOR *Ss* ASSIGNED TO EACH CONDITION

Condition	*N*	Mean	*SD*
No threat-no distraction	20	20.44	2.18
No threat-distraction	20	18.81	2.17
Threat-no distraction	20	18.83	3.45
Threat-distraction	20	19.12	3.07

counterbalancing order in each of the four combinations of threat and distraction conditions and for determining the effects of order.

RESULTS

The time, in seconds, required by each *S* to deal each deck of cards during the second of the final two series is used in the analysis. In order to reduce the variability between *Ss* attributable to differences in the sheer mechanics of dealing a deck of cards, the time required by each *S* to deal the preliminary deck (preliminary score) was subtracted from each of his time scores in the final series to provide an adjusted time score. The procedure of subtracting the preliminary score from *S*'s final time score and performing the analysis on these adjusted scores had the effect of a covariance analysis of the final time scores with time to deal the preliminary deck as the covariable. In fact, such an analysis of covariance of part of the data gave results identical to the straightforward analysis of variance of the adjusted time scores.

The means and *SD* of these preliminary scores for *Ss* in each of the four experimental conditions are given in Table 1. It can be seen from this table that the four experimental groups did not differ appreciably on these preliminary scores.

The means and *SD*s of the adjusted time scores for each range-ring pattern under each of the four experimental conditions are given in Table 2. A summary of the analysis of variance of the adjusted time scores is given in Table 3. Since the effects of order and the interactions of order with the other variables proved to be not significant, order effects are not broken down in Table 3. All effects of order are contained in the two "groups within conditions" terms in Table 3.

As can be seen from the row means in Table 2, there is a marked performance decrement under threat of shock for errors. There is a difference of 7.16 sec. between the means of the threat and nonthreat conditions which is highly significant ($P < .001$).

The over-all effects of distraction

TABLE 2

ADJUSTED TIME SCORES FOR EACH PATTERN UNDER EACH OF THE FOUR EXPERIMENTAL CONDITIONS

Condition	*N*	Range-Ring Pattern								Rows	
		A		B		C		D			
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
No threat-no distraction	20	9.12	3.70	11.09	3.57	12.35	3.34	12.75	4.31	11.33	4.00
No threat-distraction	20	12.40	4.66	13.12	3.99	14.79	4.90	15.25	4.39	13.89	4.65
Threat-no distraction	20	17.62	7.87	21.10	8.15	20.11	7.74	23.60	9.14	20.61	8.52
Threat-distraction	20	16.34	5.60	18.51	5.57	19.04	6.22	21.84	6.03	18.93	6.18
Columns	80	13.87	6.59	15.96	6.91	16.57	6.58	18.36	7.72		

TABLE 3

ANALYSIS OF VARIANCE OF THE CORRECTED TIME SCORES

Source	*df*	*MS*	*F*
Design patterns (P)	3	274.27	57.38*
Shock threat (Sh)	1	4104.58	31.92*
Distraction (D)	1	16.05	.12
P × Sh	3	33.00	6.90*
P × D	3	6.05	1.27
D × Sh	1	359.02	2.79
P × D × Sh	3	1.29	.27
Groups within conditions	12	134.84	1.05
P × Groups within conditions	36	4.49	.94
Ss within groups	64	128.57	
P × *Ss* within groups	192	4.78	
Total	319		

* $P < .001$.

proved to be not significant, there being a difference of only .44 sec. between the means of the distraction and nondistraction conditions. However, it will be noted in the row means of Table 2 that distraction affected performance differently in the threat and nonthreat conditions. In the nonthreat condition, distraction produced a performance decrement which is indicated by an increase of 2.56 sec. in the mean score. In the threat condition, however, distraction produced a performance increment which is indicated by a decrease of 1.68 sec. in the mean score. In the analysis (Table 3) this interaction of the Distraction and Threat conditions is not significant ($P = .10$).

The over-all effect of the different range-ring patterns is highly significant. The t tests indicate a difference between all adjacent patterns beyond the .001 level of significance, except for Patterns B and C, where $.05 < P < .10$. The critical consideration for the hypotheses investigated in the present experiment is whether or not the differences due to patterns are the same for the four experimental conditions. It can be seen from Table 2 that these differences between pattern means are, indeed, not the same for all experimental conditions. The effect of threat of shock was to increase the differences between pattern means. For example, the difference between the means of the best (A) and the worst (D) patterns, in terms of performance, under the nonthreat conditions is 3.24 sec., while the difference between the means of these same patterns under threat of shock is 5.74 sec. It should be pointed out that, under the threat of shock condition, there is an increase in the differences between the means of all adjacent patterns except for Patterns B and C where there is a decrease in the difference. The over-all interaction of patterns and threat of shock, as shown in Table 3, is significant beyond the .001 level. For all adjacent patterns, except B and C, this interaction is significant beyond the .01 level.

Distraction, on the other hand, had no significant effect on the differences between pattern means, as indicated by the nonsignificant interaction between patterns and distraction shown in Table 3.

As might be expected, *Ss* made occasional errors, even though they were instructed to avoid doing so. The total number of errors made with

TABLE 4

TOTAL ERRORS FOR EACH PATTERN UNDER EACH CONDITION

Conditions	*N*	Range-Ring Pattern				Total
		A	B	C	D	
No threat-no distraction	20	4	16	16	27	63
No threat-distraction	20	13	7	13	20	53
Threat-no distraction	20	3	9	1	2	15
Threat-distraction	20	2	2	2	5	11
Total	80	22	34	32	54	142

each range-ring pattern for each condition is given in Table 4. As seen in this table, the total number of errors is small, representing less than 1.4 errors for each 100 identifications. The only consistent effect is the decrease in the number of errors under threat of shock. Also, while there is some tendency for more errors to occur with the poorer patterns in the nonthreat condition, the errors are fairly constant from pattern to pattern under threat of shock for errors.

Discussion

The significant main effects of range-ring patterns and threat of shock on performance were, of course, expected and serve only to confirm similar findings already reported in the literature. The effects of distraction, however, were not anticipated. The simple effects of distraction on performance were different under the threat and nonthreat conditions. In fact, there is an inverse relationship between these simple effects which accounts for the nonsignificance of the main effect of distraction in this experiment. In the nonthreat condition, distraction produced an over-all performance decrement; in the threat condition, however, the introduction of distraction was accompanied by a performance increment. Though the data reported here do not demonstrate this interaction of distraction and threat at better than the .10 level of significance, related distraction effects which have been reported by other *Es* suggest that this interaction is probably a real effect rather than a chance occurrence. Auble and Britton (1958) found that a high anxiety group performed significantly better under a noise distraction condition than under quiet conditions while a low anxiety group performed better under quiet conditions. Angelino and Mech (1955) reported a similar inverse relationship between adjustment and performance under a noise distraction condition. The results of the present experiment suggest that distraction may have functioned in the threat condition so as to reduce the effects of threat. A possible explanation for this interaction effect is suggested by Lazarus, Deese, and Osler (1952). They suggest that the compelling preoccupation with consequences of failure may be disrupting to performance, especially on tasks requiring concentration. If we can assume that preoccupation with the consequences of making an error in this experiment (i.e., receiving a shock) was disruptive to performance, we may go a step further and suggest that the addition of distracting stimuli served to distract from this preoccupation and thereby reduce the effects of threat of shock.

The highly significant interaction between patterns and threat of shock, supports the hypothesis that the differences in performance obtained with different task designs are greater under threat of shock than under nonthreatening conditions. It is, of course, not surprising that the time scores should increase under the condition of threat of shock for errors. The important point is that there is a differential increase in time depending on the ring pattern employed and that the increase in time is greater for the worst pattern than for the best pattern. If we may assume, as was suggested earlier, that some task designs are inherently more stressful than others, then we may regard the differential effect of threat on the performance obtained with the different ring patterns used in this study as the result of the combined effects of two sources of stress. In other words, the increment of stress provided by threat of shock has more effect on performance for the pattern whose inherent level of stress is greater. The suggestion of Lazarus, Deese, and Osler (1953), which was mentioned earlier, seems to apply here as well. They suggest that the compelling preoccupation with consequences of failure may be disrupting for tasks requiring concentration while having lesser effect on tasks demanding little or no concentration. The findings

of the present experiment are consistent with this suggestion, since *Ss*' comments would indicate that those in the threat group were, indeed, preoccupied with the consequences of making an error, and it seems reasonable that the more demanding patterns require more concentration.

A comparison of Tables 2 and 4 suggests a further explanation for the interaction obtained between patterns and the threat variable. In the nonthreat condition performance became progressively poorer from Patterns A to D as measured by either time or error scores. However, under threat of shock, errors are not only markedly reduced but they are fairly constant from pattern to pattern. It seems, therefore, that in the situation where errors do not appear to be "costly," some *Ss* may be willing to tolerate a certain probability of error in order to save time. Such errors are more evident with the more difficult patterns. However, when errors are made "costly," *Ss* attempt to avoid errors even at the expense of greatly increased time. This has the effect of stabilizing errors at a minimum level. Since without stress, the probability of making an error is greater on the more difficult patterns, the nearly complete reduction of errors should increase the time scores more for the more difficult patterns.

The second hypothesis, that differences in performance obtained with different task designs are greater under distraction than under nondistracting conditions, is not supported by the findings. The absence of an interaction between patterns and distraction indicates that distraction, as induced in this experiment, failed to influence differences in performance obtained with different patterns. Since this hypothesis seemed reasonable to begin with, one might speculate as to why it was not supported by these findings. Perhaps distraction was regarded by *Ss* as "something to beat," and accordingly may have affected the level of effort of the distraction group. Others (Lazarus et al., 1953) have pointed out that the effects of distraction in laboratory experiments are apt to be confounded with the effects of motivation. Perhaps an increase in level of effort had the effect of decreasing the differences in performance obtained with different task designs. An increase in level of effort, then, might very well have masked the effects of distraction on the differences in performance obtained with the range-ring patterns in the present experiment.

In addition to the above considerations, the results of this experiment have implications for other types of psychological research. It shows that differences in task design may, in fact, be much more effective in producing differences in performance than one would be led to expect from the usual experiment. Many tasks are in fact performed under conditions of threat or "anxiety"—certainly this is true for most military tasks. Under these conditions it is reasonable to expect task design to be a more critical factor than is often assumed.

Summary

The purpose was to investigate the effects of threat of electric shock, aural distraction, and task design on performance. Eighty male *Ss* dealt four decks of 32 cards, each deck consisting of a different design pattern of eight range-rings. The time to deal each deck and the number of errors made in identifying a marked ring on each card were recorded. Twenty *Ss* were assigned to one of four experimental groups which involved different combinations of two levels of threat and two levels of distraction. Threat was induced by attaching an electrode to *S* and informing him that he would be shocked for each error. Distraction was produced by presenting a series of random numbers aurally to *S* through a set of earphones.

Performance differences obtained with the different range-ring patterns and measured in terms of time to deal a deck of cards were greater under the threat condition than under the nonthreat condition. Distraction had no significant effect on the differences in performance obtained with the different range-ring patterns.

There was some indication ($P = .10$) of an interaction between threat and distraction conditions. In the nonthreat condition, distraction produced a performance decrement, while in the threat condition, distrac-

tion was accompanied by a performance increment.

The total number of errors amounted to less than 1.4 errors for each 100 identifications. However, there were significantly fewer errors in the threat condition than in the nonthreat condition. In the nonthreat condition, there was a tendency to make more errors with the poorer range-ring patterns. In the threat condition, the errors were fairly constant from pattern to pattern.

REFERENCES

ANGELINO, H., & MECH, E. V. Factors influencing routine performance under noise: II. An exploratory analysis of the influence of "adjustment." *J. Psychol.*, 1955, **40**, 397–401.

AUBLE, D., & BRITTON, N. Anxiety as a factor influencing routine performance under auditory stimuli. *J. gen Psychol.*, 1958, **58**, 111–114.

CHAPANIS, A. Accuracy of oral transmission of target information. Systems Res. Field Lab. (Jamestown), Memo. Rep. 1954.

CONRAD, R. Speed and load stress in a sensory-motor skill. Appl. Psychol. Res. Unit (Cambridge), 1950, APU 134/50.

GARNER, W. R., SALTZMAN, D. C., & SALTZMAN, I. J. Some design factors affecting the speed of identification of range rings on polar coordinate displays. Johns Hopkins Univer. Psychol. Lab. (ICR), Tech. Rep., 1949, SDC 166–I–95.

GATES, G. S., & RISSLAND, L. Q. The effect of encouragement and discouragement upon performance. *J. educ. Psychol.*, 1923, **14**, 21–26.

HURLOCK, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, **11**, No. 71.

LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MACKWORTH, N. H. Some recent studies of human stress from a marine and naval viewpoint. *Inst. of Marine Engrs. Trans.*, London, 1952, Vol. 64, No. 6.

VERVILLE, E. The effect of emotion and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133–145.

(Received August 18, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

SOME EFFECTS OF INTRODUCING AND WITHDRAWING KNOWLEDGE OF RESULTS EARLY AND LATE IN PRACTICE [1]

EDWARD A. BILODEAU, INA McD. BILODEAU, AND DONALD A. SCHUMSKY

Tulane University

An earlier study by Bilodeau and Bilodeau (1958) reported the effect of providing schedules of knowledge of results (KR) according to fixed ratio. It was shown that learning depended upon the absolute frequency of KR, but not upon the relative frequency. That is, whatever the distribution of KR, non-KR trials neither hindered nor facilitated the learning produced by KR trials. The present study also deals with the relevance of KR for learning, but here KR and non-KR trials were administered in blocks. Such blocks were introduced either early or late in practice.

Because of the results of the previous study, it was expected that performance would improve with KR and deteriorate after the withdrawal of KR. The major expectation, however, was that non-KR trials would have no effect on learning when KR was eventually introduced. Actually, the issues involved here are analogous to those of the latent learning, reward context, and the experimental design was structured accordingly. In the latter part of the experiment all *S*s were tested with KR. Prior to this, one group practiced without KR, another with KR, and two others with a different number of KR trials followed by trials without KR.

A secondary purpose of the experiment was to provide observations on the rate and level of response deterioration as KR was withdrawn from *S*s of diverse proficiency.

METHOD

Subjects and apparatus.—One hundred and sixty naive male *S*s were assigned in equal number and without known bias to each of four conditions of practice on the Manual Lever apparatus. A complete description of the device was published by Bilodeau and Ferguson (1953).

The *S*'s task was to learn to displace the lever by a certain amount. Unknown to *S*, a displacement of 33° of arc constituted a perfect pull of the lever. Whenever KR was given, it followed the pull by about 5 sec.; "10 units high" and "8 units low" are examples of KR. The task, then, was to learn to pull the lever by an amount which would minimize the error reported verbally by *E*.

A 20-lb. pull was required to move the lever. Thus, proprioceptive cues were strong as compared with a line drawing task.

Procedure.—For each of four groups, 24 pulls were required. The last five pulls were always followed by KR. For Groups 0, 2, 6, and 19, the first 0, 2, 6, and 19 trials, respectively, were also with KR. The remaining trials were not.

All 24 trials were run with a 20-sec. trial cycle. A "ready" signal preceded the "pull" signal by 3 sec. The response was made within 3–4 sec., and KR was given 5 sec. after response completion. About 8 sec. remained until the next warning signal. The *S* released the lever on completion of the pull, returned his arm to his side and waited for the next ready signal.

RESULTS [2]

The mean absolute errors (actual response less 33°) are plotted in Fig. 1.

[1] The data were collected at the Air Force Personnel and Training Research Center, San Antonio, Texas. The costs of preparing and publishing the manuscript were borne by the Graduate Research Council of Tulane University.

[2] Two tables of data have been deposited with the American Documentation Institute.

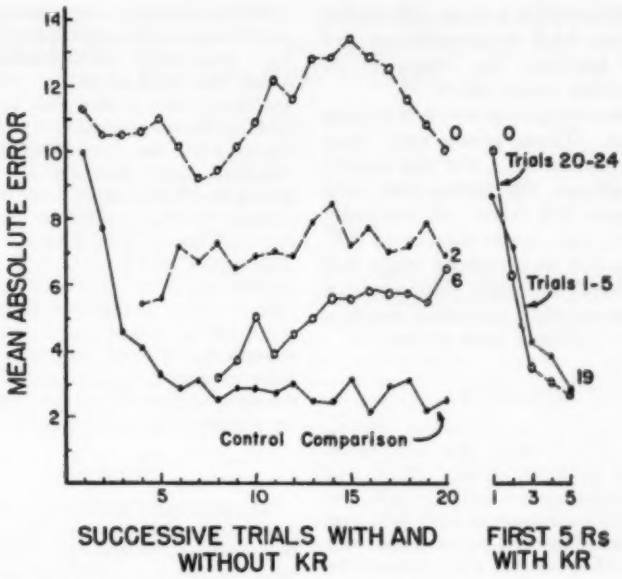

FIG. 1. Mean absolute error in positioning a Manual Lever to 33° on successive trials with and without KR (see text).

The figure is divided into two independent parts. In the first of these parts, the curve labeled Control Comparison shows the effects of 19 successive KRs. The Control Comparison represents the errors of three groups (2, 6, and 19) averaged whenever KR was in effect. The other curves on the left side of Fig. 1 show no particular trend in the error when KR is entirely absent (see 0), and increasing error trends after KR is withdrawn (see 2 and 6). The increasing error trends of Groups 2 and 6 are marked, though they do not suggest an imminent regression to the level of the initial error. The performance of Group 6 certainly deteriorates no more slowly than that of Group 2 despite the fact that Group 6 has had four more KRs. Groups 2 and 6 do not appear sufficient for a firm conclusion on deterioration as a function of the number of previous KRs. A group with KR withdrawn after 20 or so trials would have been a useful addition to the experimental design.

The second part of Fig. 1 presents a test of the latent effect of repeated responding in the absence of KR. Here, the last five trials of Group 0 represent the *first* trials with KR, after a series of 19 trials without KR. Group 19 served as a control, since their first five KR trials provided their first experience with the apparatus. The mean squares of neither Groups nor Groups by Trials reached the .05 level of significance. Thus, introducing KR immediately

Copies may be secured by citing Document Number 5957 and by remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress, Washington 25, D. C.

or late in practice had no differential effect upon level of performance and rate of learning, i.e., there is no demonstrable latent effect.

The mean *algebraic* errors were also examined. These errors were near zero for Group 19. For the experimental groups, the errors were near zero during KR trials and markedly positive in sign in the absence of KR. Not only did *S*s overshoot when KR was omitted, but they also overshot by amounts that increased steadily the longer KR had been absent.

Discussion

The results were consistent with a previous study where KR was given according to fixed ratios. There was (*a*) no improvement without KR, (*b*) progressive improvement with KR, and (*c*) deterioration of response proficiency after the withdrawal of KR. The curves of Fig. 1 effectively summarized the classical properties of KR. Furthermore, it was shown that an initial series of trials without KR did not serve to acquaint *S* with properties of the lever system that could be gainfully employed (positive transfer) with the introduction of KR after the twentieth response. However, and lamentably so, a failure to produce latent learning is not to say that latent learning cannot be produced (though we may not know how).

A few KR trials (2 or 6) introduced at the very beginning of practice were effective in producing relatively long-term response modification, though response deterioration did set in immediately upon the withdrawal of KR. The deterioration, marked in quantity and positive in sign, nonetheless remained substantially below the level of errors produced when KR was entirely absent. The positive effect of a single KR trial is probably much greater (and more lasting) than the negative effect of a non-KR trial.

Dees and Grindley (1951), and others, have also observed increasing amplitude of response upon withdrawal of KR. A naive application of Hull's I_R theory (1943) to deterioration phenomena would predict *decreasing* amplitude of response, e.g., progressive undershooting would follow the build-up of I_R. Perhaps, it would be better to state that I_R interacts with or damps proprioceptive feedback. Thus, with the accumulation of I_R, responses must progressively lengthen during non-KR trials in order to produce internal feedback sufficient on Trial n to match feedback on Trial $n - 1$. The assumption here is that S is matching proprioceptive feedbacks during trials without KR. If this were done without the accumulation of I_R, a series of responses would be of uniform amplitude. With increasing I_R the amplitude of the response might well rise in compensation for a decreasing proprioceptive feedback.

Summary

This experiment deals with the late introduction and removal of knowledge of results. Its purpose was, by showing in a single experiment the effects of basic manipulations, to reassert the powerful relationships between response and the presence or absence of KR. The task was lever-displacing; the stimulus (KR) was the amount and direction of the reported error.

The experiment showed (*a*) no improvement without KR, (*b*) progressive improvement with KR, and (*c*) response deterioration after the withdrawal of KR. Further, an early series of trials without KR had no latent effect on the learning shown when KR was eventually introduced.

References

Bilodeau, E. A., & Bilodeau, I. McD. Variable frequency of knowledge of results and the learning of a simple skill. *J. exp. Psychol.*, 1958, **55**, 379–383.

Bilodeau, E. A., & Ferguson, T. G. A device for presenting knowledge of results as a variable function of the magnitude of the response. *Amer. J. Psychol.*, 1953, **66**, 483–487.

Dees, V., & Grindley, G. C. The effect of knowledge of results on learning and performance: IV. The direction of the error in very simple skills. *Quart. J. exp. Psychol.*, 1951, **3**, 36–42.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

(Received August 21, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

DIFFERENTIAL EFFECTS OF CONTINUOUS EXTINCTION AND DISCRIMINATION TRAINING ON THE GENERALIZATION GRADIENT [1]

WERNER K. HONIG

Denison University

DAVID R. THOMAS [2] AND NORMAN GUTTMAN

Duke University

A basic assumption in the traditional stimulus-response analysis of discrimination learning is that excitatory and inhibitory response tendencies are independently developed to the positive (S^+) and negative (S^-) stimuli as a result of the action of reinforcement and nonreinforcement. The only "interaction" occurring between these stimuli results from the generation of gradients of excitation and inhibitation which have S^+ and S^- as their respective maxima, and which summate algebraically on the stimulus dimension on which S^+ and S^- are differentiated. This analysis, first proposed by Spence (1937) to account for transposition, permits the prediction that after discrimination training, the resulting generalization gradient has its maximum not at S^+ but at a previously neutral stimulus value displaced away from the S^-.

To support the assumption of gradients of excitation and inhibition, Spence cites the work of Pavlov (1927) and of Bass and Hull (1934) in the area of classical conditioning. More recently, Guttman and Kalish (1956) have made a reliable determination of generalization gradients along the spectral dimension in an operant conditioning situation. Honig (1956) was able to establish a gradient of extinction after training a response to a number of stimulus values and extinguishing on one value. In unpublished research, Jenkins [3] obtained a gradient of extinction on the dimension of auditory frequency, after training a discrimination where silence served as S^+ and a tone as S^-. These experiments give support to Spence's assumptions about the existence of gradients of excitation and inhibition, although the forms of the gradients are not always similar to those postulated. The question remains whether the achievement of a discrimination is merely an additive combination of the processes of excitation and inhibition at different points on the stimulus continuum, or whether some more complex interaction is involved.

Hanson (1956, 1957) has investigated the spectral generalization gradient after the establishment of a successive discrimination. He did indeed obtain a displacement of the maximal value, or "peak shift," in the direction predicted by Spence. However, he found considerable difficulty in attempting to derive his empirical gradients from the summation of postulated excitatory and inhibitory gradients. His discussion is too complex to be reviewed here, but one point is particularly relevant to the present study: any postulated inhibitory gradients accounting for his data would have to have maxima at the S^+ rather than the S^- value, a determination that is certainly contrary to Spence's analysis. Hanson concludes that the best description of the effect of discrimination training is a "relocation" of the primary generalization gradient along the stimulus continuum. The simple summation of independently generated gradients of excitation and inhibition cannot account for this "relocation"; the development of a discrimination appears to involve additional processes that are *sui generis*, and

[1] This study was supported by Grant MH-1002 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Deborah Nickerson and Robert Langston for assistance in running the *Ss*. A modified version of this paper was presented at the meetings of APA, New York, 1957.

[2] Now at Kent State University.

[3] Personal communication from Herbert M. Jenkins, August, 1958.

not accounted for in terms of the combination of effects of simple acquisition and extinction.

This problem may be approached in another way: If the additive effects of excitation and inhibition are responsible for the effects of discrimination training, extinction at some point on the generalization gradient should have the same effect on the gradient as discrimination training. While there may be some differences, particularly in over-all response level, due to the continued presentation of reinforcements in successive discrimination training, some tendency to a peak shift should be observed in both cases. The present experiment was designed to test this hypothesis. After being trained to respond to one stimulus (S^R), pigeons were given massed, or continuous extinction at another value (S^E) lying on the generalization gradient. The effects of the extinction could be estimated by comparing the subsequent generalization gradient with a control gradient obtained when no extinction was given. If a gradient of extinction is generated after continuous extinction at S^E, it should be observed in such a comparison. After the postextinction gradient was obtained, the *S*s were given successive discrimination training with S^R and S^E serving as S^+ and S^-, respectively, so that the effects of this training could be compared with those of continuous extinction.

Method

The study was carried out in two separate experiments identical in procedure but differing slightly with regard to breed of *S*s and brightness equalization of the stimuli. Since these differences appear to be responsible for differences in height and form of the generalization gradients, the data from the two experiments have not been combined, although they are presented and discussed together.

Subjects.—Eighteen homing pigeons were used in Exp. I. Twenty-two White Carneau pigeons were used in Exp. II.

Apparatus.—An automatic key-pecking apparatus similar to that described by Guttman and Kalish (1956) was used in both experiments. The light source for the illumination of the key was a Bausch and Lomb monochromator, which employs a diffraction grating for the dispersion of white light into the spectral continuum. The band width of the spectral stimuli used in the present study was about 16 mμ. In Exp. I, the stimuli were equated for apparent brightness by means of a neutral density optical wedge. A human observer made the equations with the aid of a Macbeth illuminometer. Blough's finding (1957) that the pigeon's luminosity function is similar to that of man provides justification for this procedure. In Exp. II, brightness values were not equated.

Procedure.—Preliminary and variable interval training: The *S*s were reduced by starvation to between 75 and 80% of their free-feeding body weight. They were magazine-trained and conditioned by successive approximations to peck at the key illuminated by a light of 550 mμ. After one session of 50 continuous reinforcements, variable interval (VI) training was begun. VI training extended over 10 daily sessions, each consisting of 30, 1-min. "stimulus-on" periods in which the 550 mμ light was presented alternately with 10-sec. periods of blackout. The VI schedule had a mean inter-reinforcement interval of about 1 min. The programmer did not operate during the blackout periods, and no reinforcements could be obtained at that time.

Continuous extinction and the first generalization test: On the day following the last VI session, continuous extinction on 570 mμ was given by presenting that stimulus without reinforcement for 1-min. periods separated by 10-sec. blackout intervals. Group 1 (six *S*s, Exp. I; eight *S*s, Exp. II) received 20 min. of extinction, while Group 2 (six *S*s, Exp. I; eight *S*s, Exp. II) received 40 min. The control Group (six *S*s, both experiments) received no extinction. On the following day, each *S* was tested for generalization in extinction. Eleven stimuli were used, ranging from 490 to 610 mμ in 10 mμ intervals, with the omission of 500 and 600 mμ. Each test consisted of 12 blocks of stimulus presentations in which each stimulus occurred once. The 11 stimuli were presented in a random order within each block. Each presentation was for 30 sec. and was followed by a 10-sec. blackout period to allow for the changing of the stimulus and the recording of data. Since the control group received no extinction, its generalization test was given on the day following the completion of VI training.

Discrimination training and the second generalization test: On the day after the first test, each *S* from Groups 1 and 2 received 10 continuous reinforcements with the 550 mμ stimulus in order to reinstate the conditioned operant. Successive discrimination training started immediately thereafter. The latter was carried out for 80 min. in all: 30 min. on Days 1 and 2, and 20 min. on Day 3. Responding to 550 mμ, the positive stimulus, was reinforced according to the VI schedule described above. Responding to 570 mμ, the negative stimulus, was never reinforced. Each stimulus was presented for 1 min. periods followed by 10 sec. of blackout. Within each block of 10 periods, S^+ and S^- were each presented five times. The order of presentation was randomized from block to block according to a Gellerman series (Stevens, 1951) no stimulus being presented for more than three successive periods. One *S* from Group 2 of Exp. II died before the completion of this phase of the research, reducing the *N* for that group to five.

Unlike the continuous extinction procedure, successive discrimination training was carried out for an equal amount of time for all *S*s, with 40 min. of presentation of each stimulus. A few *S*s stopped responding entirely during part of the scheduled discrimination training procedure. If the cessation of responding continued for 10 or more successive periods, these presentations were repeated at the end of the third training session in order to insure that all *S*s received equal amounts of "active" training.

Results

The effects of continuous extinction.— The postextinction gradients (PEGs) and control gradients are presented in Fig. 1 in terms of the mean number of responses obtained at each stimulus value on the first generalization test. In order to estimate the effects of the extinction procedure, the PEGs have also been subtracted from the control gradients, the differences being shown below the line of zero response level by curves which may be called "gradients of decrement." From inspection of Fig. 1 it is apparent that (*a*) the PEGs from Exp. I are very similar, (*b*) they are considerably lower than the control gradients, (*c*) the reduction indicated by the gradient of decrement is a general one and appears to be roughly proportional to the height of the control gradient at each point. The decrement is not centered around 570 mμ; it is smaller, in fact, at that point than at 550 and 560 mμ, and, for the 40-min. group, at 580 mμ. In Exp. II there appears to be a difference in the form of the PEGs, that for the 20-min. group being somewhat steeper.

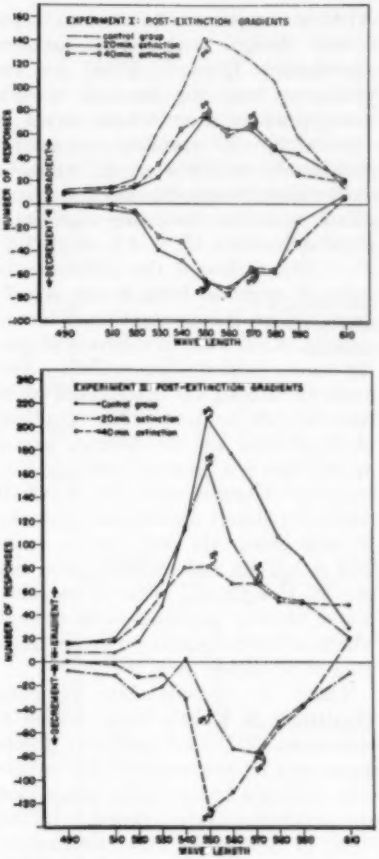

FIG. 1. Postextinction and control gradients from Exp. I (above) and Exp. II (below).

This is supported by a two-factor mixed design analysis of variance (Lindquist Type I, 1953) on the gradients from the 20- and 40-min. groups, where wave length serves as the "within-*Ss*" variable, and extinction time as the "between-*Ss*" variable. A significant interaction between treatment variables indicates significant slope differences ($F = 2.2$; 10, 125 *df*, $P < .05$), although the difference in over-all response level is not significant. Since the assumption of homogeneity of variance at different stimulus values is of dubious validity, this analysis should be interpreted with considerable caution. The gradient of decrement for the 40-min. group again shows a general reduction in response strength, with the maximal reduction clearly at 550 mμ. For the 20-min. group, the reduction is maximal at 570 mμ and decreases to either side of this value. This is the only one of the four gradients of decrement which actually appears to be centered around S^E rather than S^R.

There is considerable evidence (Guttman & Kalish, 1956; Kalish & Guttman, 1957) that gradients differing in over-all response level but otherwise obtained under similar conditions are multiplicatively related. If the effect of continuous extinction represents no more than a general reduction in response level, a multiplicative relationship should be observed between the PEGs and the respective control gradients. In such a case, the percentage obtained at each stimulus of the total number of responses from the entire gradient should be identical for the PEGs and control gradients. These percentages are plotted in Fig. 2. It is clear that in Exp. I they are similar for all groups. In Exp. II, on the other hand, there appear to be systematic differences between the percentage gradients. That of the

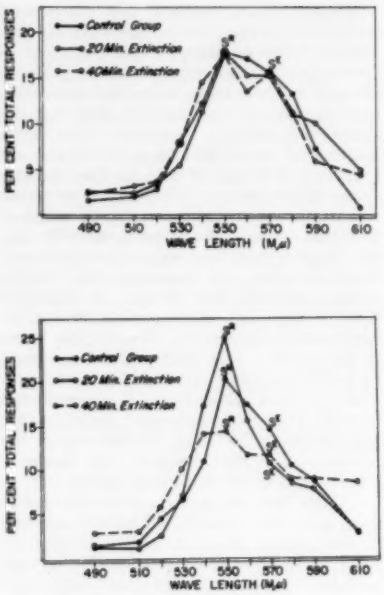

FIG. 2. Postextinction and control gradients from Exp. I (above) and Exp. II (below) plotted in terms of percentage of total responses at each stimulus value.

40-min. group is flatter than that of the control group, while that of the 20-min. group shows a higher percentage at 540 and 550 mμ, and a lower percentage from 560 to 590 mμ, with the largest difference occurring at the S^E. An analysis of variance similar to that described above on the three percentage gradients from Exp. II shows a significant interaction between the treatment variables ($F = 2.94$; 20, 190 *df*, $P < .01$). Since the slope differences between the gradients of the 20-min. group and the control group are of particular interest, a separate analysis was done with these two gradients. No significant interaction was obtained, ($F = 1.36$; 10, 120 *df*) so that even the apparently specific effect in the

TABLE 1

MEANS OF THE PEGS AND PDGS WITH THE DIFFERENCE BETWEEN EACH EXPERIMENTAL GRADIENT MEAN AND THE RESPECTIVE CONTROL GRADIENT MEAN

Group	Postextinction		Postdiscrimination	
	Mean PEG	Exp.-Control Difference	Mean PDG	Exp.-Control Difference
Experiment I				
20-min.	559.3 mμ	−1.7	549.3 mμ	8.3
40-min.	556.1	1.5	549.9	7.7
Control	557.6			
Experiment II				
20-min.	555.9	1.6	545.1	12.4
40-min.	556.8	.7	542.2	15.3
Control	557.5			

area of S^E is put into doubt; the greatest part of the interaction is presumably contributed by the flattened gradient of the 40-min. group.

A measure which is particularly sensitive to changes in the form of the gradient is the gradient mean, as it reflects the response level at all test values. If the present gradients are multiplicatively related, their means will be the same; a specific decrement in the region of the S^E, however, will be reflected by a mean shifted toward the lower wave lengths. The gradient means are presented in Table 1 together with the differences between each experimental gradient mean and the relevant control gradient mean. The means of the PEGs reflect little, if any, change from the control gradients, the largest decrease being 1.6 mμ. In one case there is even an increase of 1.7 mμ.

The effects of discrimination training.—The postdiscrimination gradients (PDGs) are presented in Fig. 3 in terms of the mean number of responses obtained at each stimulus value on the second generalization test. Compared to the control gradients, the PDGs show a suppression of responding in both experiments at values above 550 mμ. At 550 and 540 mμ, the response level is higher after discrimination training. In Exp. II, the mode of the PDG is at 540 mμ

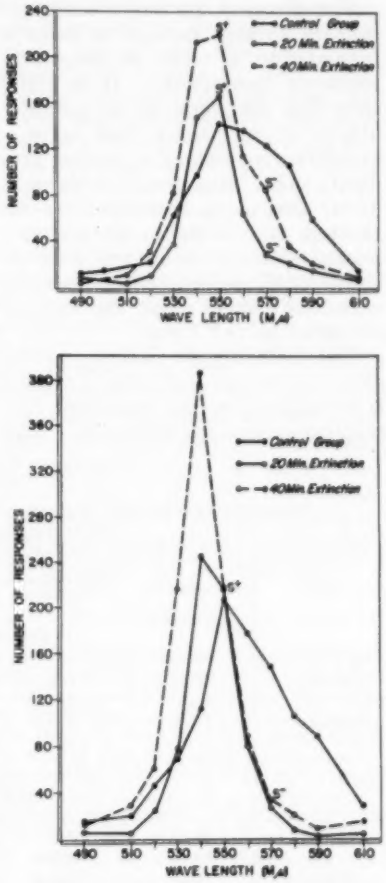

FIG. 3. Postdiscrimination and control gradients from Exp. I (above) and Exp. II (below).

for both groups; this is particularly striking for the 40-min. group. In Exp. I, the mode remains at 550 mμ, but the value at 540 mμ is high enough to suggest a "truncated" peak between 540 and 550 mμ, which may be estimated by a method of linear extrapolation suggested by Hanson (1956). This procedure reveals an estimated peak at 546 mμ for the 20-min. group and 545 mμ, for the 40-min. group.

The effects of discrimination training are striking enough to make a presentation in terms of percentage gradients unnecessary. It is quite clear that the gradients are radically altered in a manner that is nonmultiplicative with the control gradient. The "relocation" of the gradient, similar to that obtained by Hanson, is reflected in the gradient means, which can be seen in Table 1. Every PDG mean is lower than the respective control mean, the smallest difference being 7.7 mμ.

The results suggest that any specific effect on the generalization gradient is contingent on the successive discrimination training procedure. The possibility remains, however, that the continuous extinction procedure had some subtle effect, if not on the PEG, then possibly on subsequent discrimination learning or on the PDG. In that case, we might expect systematic differences between the gradient means or some measure of discrimination learning as a function either of the time spent in continuous extinction, or the number of responses emitted under that condition. That the time in continuous extinction is not related to differences in the gradient means, either before or after discrimination training, is clear from Table 1. The differences are in some cases larger for the 40-min. group, and in some cases for the 20-min. group. To examine the relationship between the number of responses in continuous extinction and these measures, each experimental group was divided in half, separating the high and low responding *Ss* in continuous extinction. The mean number of responses obtained under that condition is presented in Table 2 for each group, together with the means of the individual

TABLE 2

VARIOUS MEASURES FOR *Ss* RESPONDING AT HIGH AND LOW RATES DURING CONTINUOUS EXTINCTION

Group	Response Rate	Mean Responses in Extinction	Mean PEG Mean	Mean PDG Mean	Discrimination Index
Experiment I[a]					
20-min.	High	650.0	558.7 mμ	549.2 mμ	.47
	Low	364.7	559.8	547.2	.06
40-min.	High	755.5	556.7	547.0	.27
	Low	423.5	555.1	553.7	.55
Experiment II					
20-min.	High	912.2	560.1	546.0	.15
	Low	618.5	550.8	544.0	.15
40-min.	High	1832.2	557.6	542.2	.28
	Low	723.0	555.6	540.1	.20

[a] Of the five *Ss* in the 40-min. group, the one giving the median rate was omitted from the analysis.

PEGs and PDGs as well as an index of discrimination learning, which is the ratio of the responses to S^- to those to S^+ in the last 10 min. of training. It might be expected that a greater number of responses in extinction would be correlated with a larger shift in the mean of the gradient and a lower discrimination index. The pairs of measures showing differences in the expected direction are underlined; they amount only to one-quarter of the total.

Discussion

The foregoing analyses lead to the conclusion that the effects of continuous extinction on the generalization gradient bear little or no relation to the S^E, and that they do not resemble the effects of successive discrimination training. The only reliable effect of the former that we have observed is a general reduction in response level of the postextinction tests. With this in view, it appears fruitless to attempt an explanation of the "relocation" of responses after discrimination training by means of the interaction of separate gradients of excitation and extinction. The changes generated by discrimination training appear at this point to be *sui generis*, and they will have to remain so until a more fruitful approach to their explanation can be suggested.

Such an approach may be represented by the study of variables that are intrinsic to the discrimination situation. The amount of shift of the mean of the gradient would serve as an important dependent variable. Various differences in training conditions may serve as independent variables; by an examination of each, we may be able to deduce those factors in discrimination training that are important for the occurrence of the peak shift. Honig (1958), for example, has shown that simultaneous discrimination training has little effect on the gradient, when compared to successive, and produces no peak shift. Hanson (1956) has reported that the peak shift varies as a negative function of the CS^+ CS^- difference. One of the authors (Thomas) has evidence from an unpublished study that the amount of training may be the vehicle through which this variable of stimulus difference operates. In his study, two groups of *Ss* showed identical peak shift after the same amount of training on two problems one of which involved a CS^+, CS^- difference more than twice the other. In the same study, Thomas showed that the peak shift occurs well before the achievement of the discrimination criterion. The present experiment also demonstrates considerable peak shift after an incomplete discrimination, as does an unpublished study by Guttman in which a discrimination was made between stimuli associated with two different reinforcement schedules, one of which was more favorable than the other.

Many basic questions about operant discrimination and the postdiscrimination generalization gradient remain unanswered. What is the effect of the proportion of negative to positive periods in the successive discrimination situation? What is the effect of the length of presentation periods? Of what influence is the reinforcement schedule associated with the positive stimulus; with the negative? These examples may suffice to indicate the kinds of questions that must be answered before the effects of discrimination training on the generalization gradient may be understood.

Summary

The study was designed to investigate the effect of continuous extinction at a stimulus lying on a generalization gradient generated by reinforcement at another stimulus value. The effects of this procedure were compared to those of presenting successive discrimination training with the CS as the positive stimulus and the stimulus at which continuous extinction was given as the negative. It was found that continuous extinction has only the effect of reducing the response strength at all points on the generalization gradient in multiplicative fashion, while successive discrimination training led to a suppression of responses, specifically in the area of the

negative stimulus, and a shift both in the mean and the mode of the gradient away from that value. The conclusion was drawn that excitation and inhibition cannot independently account for the effects of successive discrimination training; the mixed presentation of positive and negative stimuli appears to be essential. Certain variables inherent to discrimination training are suggested, which, when investigated, may clarify the basis of the effects of the formation of a successive discrimination on the generalization gradient.

References

Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47–65.

Blough, D. S. Spectral sensitivity in the pigeon. *J. opt. Soc. Amer.*, 1957, **47**, 827–833.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hanson, H. M. The effects of discrimination training on stimulus generalization. Unpublished doctoral dissertation, Duke Univer., 1956.

Hanson, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, **125**, 888–889.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

Honig, W. K. Generalization of extinction in the spectral continuum. Paper read at Midwest. Psychol. Ass., Chicago, May, 1956.

Honig, W. K. Prediction of preference, transposition, and transposition-reversal from the generalization gradient. Unpublished doctoral dissertation, Duke Univer., 1958.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

Pavlov, I. P. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Trans. and ed. by G. V. Anrep. London: Oxford Univer. Press, 1927.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received August 25, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

TIME ESTIMATION AS A FUNCTION OF LEVEL OF BEHAVIOR OF SUCCESSIVE TASKS

RUTHANNE K. S. DEWOLFE AND CARL P. DUNCAN

Northwestern University

If *S*s are asked to estimate the duration of a fixed interval of time during which they are occupied with some task (including "rest"), their estimates vary inversely with certain characteristics of the task (Axel, 1924; Gilliland, Hofeld, & Eckstrand, 1946; Gulliksen, 1927; Harton, 1938; Weber, 1933; Woodrow, 1951). Axel (1924) used the phrase "level of behavior" to refer to all such characteristics of tasks (interest, difficulty, cognitive demand, etc.) that influence estimation of passage of time. Axel's study, taken together with a similar study by Gulliksen (1927), showed that such tasks as rest or tapping a pencil, considered to be at low levels of behavior, produced overestimation of a fixed interval. An analogies test or long division problems, tasks considered to be at high levels of behavior, yielded underestimation of time. An intermediate level of behavior was assumed to be represented by number canceling or mirror reading; these tasks yielded intermediate estimates.

In a recent study by Kleiser (1953), *S* first worked on some task for a fixed (standard) interval, then rested (comparison interval) until he judged as much time had elapsed as had elapsed during the standard interval. Kleiser's results with this method were essentially the same as those found by the earlier workers; the greater the "assumed cognitive demand" of the task used during the standard interval, the shorter *S*'s estimate, i.e., the shorter the comparison interval.

The present study is an extension of Kleiser's method. Tasks assumed to represent various levels of behavior are used during both the standard and the comparison intervals. It is assumed that rest, reversed alphabet printing, and anagram solving represent low, intermediate, and high levels of behavior, respectively. These three tasks are used in all nine possible combinations as first, or standard, and second, or comparison, tasks.

If the assumption about the rank order of these tasks along the dimension of level of behavior is correct, the results should be predictable. Suppose, in one extreme case, that *S* rests during the fixed, standard interval, then attempts to reproduce that interval's duration by solving anagrams during the comparison interval. Since *S* rests during the standard interval, the interval seems long to him. He now attempts to reproduce this "long" interval while solving anagrams, a task in which time seems to go quickly, i.e., is ordinarily underestimated. Therefore, his estimate, in time elaspsed during the comparison interval, should be longer than with any other combination of the present tasks. Conversely, the other extreme case, anagrams followed by rest, should produce the shortest estimates of all. Generalizing from these extreme cases, the prediction is that the higher the level of behavior of the standard task, the shorter should be the estimates, but the higher the level of behavior of the comparison task, the longer should be the estimates.

One presentation of a standard task followed by a comparison task is called one trial. Several trials are given in order to check a finding of Kleiser's that time estimates tended to vary with progressive improvement in performance on the tasks themselves. No attempt is made to predict the results except on the first trial.

METHOD

Materials.—Reversed alphabet printing was done in blue books; a separate page was used for each trial. For the anagram task, 50 five-letter anagrams were selected from the lists of Rees and Israel (1935). Thirty randomized lists of the 50 anagrams were prepared by means of a table of random numbers. Those *Ss* receiving the anagram task during the standard intervals were assigned the lists arbitrarily numbered from 1 to 15 (all *Ss* received 15 trials). The *Ss* receiving anagrams during the comparison intervals worked on Lists 16 through 30. Each list was typed lower case in two columns of 25 anagrams each on a separate sheet of paper, with the anagrams numbered consecutively from 1 to 50.

A cardboard screen on the table at which *E* and *S* sat prevented *S* from seeing any of the materials or the .01 sec. Standard Electric clock used to time both standard and comparison intervals.

Design.—The three tasks, rest, reversed alphabet printing, and anagram solving were used in all nine combinations in both the standard interval and the comparison interval. The *Ss*, 135 students in introductory psychology courses, were assigned in turn to the nine groups, 15 *Ss* per group. The duration of the standard interval was 26 sec. on each trial. There were 15 trials, each consisting of a standard and a comparison interval. The duration of the comparison interval, to the nearest whole second, was *S*'s score. The *S* did not know the duration of the standard interval, and no knowledge of results was given.

Procedure.—Each *S* was first given a pretrial with each of the tasks on which he would eventually work. For rest, the pretrial consisted of having *S* lean back in his chair, close his eyes, and relax as completely as possible for 5 sec. For anagrams, *S* solved in the pretrial as many as he could of a list of 10 practice anagrams. The pretrial for reversed alphabet printing consisted of printing the alphabet once, printing the letters upside down and from right to left.

After the pretrial, *S* was instructed as follows. He would be given a number of trials, each of which consisted of two parts. For the first part of each trial *E* would say "Go" and *S* would engage in the activity assigned for the standard interval until *E* said "Stop." For the second part of each trial, *E* would again say "Go" and *S* would engage in the activity assigned for the comparison interval. Then *E* said, "However, during the second part of the trial you will continue to rest (solve anagrams, or print the alphabet in reverse, as appropriate) only until you feel that you have rested (worked) as long as you rested (worked) during the first part of the trial when I said 'Stop.' Then you must say 'Stop.' When you feel that as much time has elapsed during the second part of the trial as elapsed during the first part of the trial when I said 'Stop,' you must say 'Stop.' You will be trying to match the two times exactly." The *S* was told not to use any kind of counting method on which to base his reproductions.

After these instructions, *S* was given instructions specific to the task or tasks he would be working on. The *Ss* who were to do reversed alphabet printing were told to work as quickly and as accurately as possible. If in a single period (half-trial) they completed the alphabet, they were to continue right on with *A* again, but at the beginning of each new period they were to start with the letter following the last letter printed in the preceding period. The *E* presented *S* with a fresh page in the blue book for each half-trial and told him the letter with which to begin. The *Ss* were instructed not to copy previous work, but to print each letter independently.

The *Ss* who were to solve anagrams were told that during each period they were to solve as many anagrams as possible without using pencil and paper. As soon as *S* had solved an anagram, he was to call out its number and correct solution, and go on to any other anagram on the sheet.

Before *each* standard or comparison interval, *E* said to *S*, "This time you will be resting (or solving anagrams, or printing the alphabet in reverse). Remember, this time it is your (my) turn to say 'Stop.' You must (I will) say 'Stop' this time." Giving this instruction and handing *S* whatever materials he needed took no more than 5 sec. between standard and comparison intervals. Time between trials was also about 5 sec.

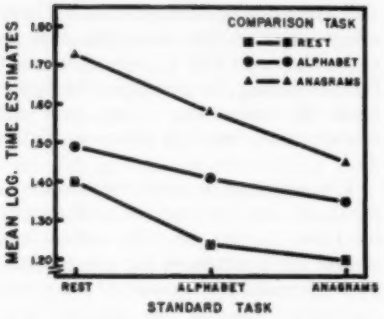

Fig. 1. Mean log time estimates on Trial 1.

Results

Trial 1.—Since an attempt was made to predict the results on Trial 1, the data for this trial will be analyzed separately. Inspection of the Trial 1 data revealed some positive skew in in each of the nine distributions. Because of this, each raw score was converted to a logarithm. Hartley's test for heterogeneity of variance among the log scores yielded $F_{max} = 2.33$, which, with 9 and 14 *df*, is not significant.

The mean log time estimates are plotted in Fig. 1. Because the precise location of the tasks along the dimension "level of behavior" is not known, the tasks have been arbitrarily spaced equidistantly from each other on the baseline. By inspection, it appears that both variables had an effect in the predicted direction; estimates decrease with increasing level of behavior of standard task and increase with increasing level of behavior of comparison task.

The analysis of variance on the means supported the trends suggested in Fig. 1. The *F* for standard task was 33.0; the *F* for comparison task was 69.6. With 2 and 126 *df*, both these values are significant at the .001 level. The interaction was not significant.

Although no attempt was made to measure the accuracy with which *Ss*, working on various tasks, can reproduce a 26-sec. interval when they know the duration of the interval, it is of interest to note, in Fig. 1, that the absolute duration was reproduced most accurately by those three groups whose comparison task was the same as their standard task. The log of 26 is 1.415; the means for the Rest-Rest, Alphabet-Alphabet, and Anagram-Anagram groups were 1.40, 1.41, and 1.45, respectively.

Trend effects.—To examine the data for changes in time estimates over the 15 trials, the 15 raw scores were combined into five blocks of three trials each. A graph (not presented here) of these five trend points for each of the nine groups suggested that there was very little change in time estimates over the 15 trials. A trend analysis on these data, to test for the effect of trials, showed that neither Trials, nor any of the interactions with Trials, was significant; time estimates did not vary with repeated experience with the two-task procedure (even though scores on both the alphabet and anagram tasks improved over trials). The scores were also fairly stable in another respect; the correlation between mean score on Trials 1–3 and mean score on Trials 13–15, based on all 135 *Ss*, was .803.

The trend analysis also yielded terms for the effects of the main variables on total raw score over all 15 trials; these terms are shown in Table 1. It can be seen that in contrast to the score on Trial 1 alone, the interaction between standard and comparison tasks was significant. Because of this, the interaction was used as error to compute the *F*s, shown in Table 1, for the main variables. Even though the *df* are only 2 and 4,

TABLE 1

EFFECT OF MAIN VARIABLES ON TOTAL RAW SCORE OVER ALL TRIALS

Source	*df*	*MS*	*F*
Standard task (S)	2	101,490.60	4.36
Comparison task (C)	2	223,557.59	9.60*
S × C	4	23,288.71	8.04***
Between *Ss* in same group	126	2,896.73	

* $P < .05$.
*** $P < .001$.

the F for the comparison task was significant at the 5% level, and the F for the standard task reach the 10% level. (With Between-*Ss*-in-same-group as error, the *F*s for standard and comparison tasks were 35.04 and 77.18, respectively.)

Since the effects of the main variables on total raw score over all trials were not the same (because of interaction) as on log score on Trial 1, the mean of each *S*'s score over all 15 trials was converted to a log to provide greater comparability with Trial 1 data. The F_{max} on the variances of these log scores was 19.35, significant well beyond the 1% level, so the results must be interpreted conservatively. The analysis of these log scores showed that the interaction between the main variables was significant ($F = 3.31$), although only at the 5% level in this case. With the interaction as error, both main variables were significant; $F = 8.78$ ($P < .05$) for standard task, $F = 21.56$ ($P < .01$) for comparison task. (With Within-groups as error, $F = 29.04$ for standard task, $F = 71.32$ for comparison task.) Despite heterogeneity of variance, on the whole it appears that on mean score over all 15 trials there was a significant interaction between standard and comparison task variables, but that over and above the interaction, the variables still exerted significant effects. These effects were in the same direction as those shown in Fig. 1, as may be seen by comparing the points in the figure with the mean log score over all trials, shown for each group in Table 2.

The significant interaction among standard and comparison task means in Table 2, i.e., the interaction revealed by analysis of log mean score over all trials, is the result of the same patterning among the means that occurred on Trial 1 (where the interaction though not significant, was greater than zero). Both in Fig. 1 and in Table 2, when rest was the standard task, the largest difference occurred between alphabet and anagrams as comparison tasks, but when anagrams were the standard task, the largest difference occurred between rest and alphabet as comparison tasks. In addition, there is a reversal (not significant) in Table 2 between the alphabet and anagram means for standard task when rest was the comparison task.

Table 2 shows that the means of the three groups whose standard and comparison tasks were the same, increase as one proceeds from Rest-Rest to Anagram-Anagram conditions, as was also the case on Trial 1. This increase suggests that the comparison task had a more powerful influence on estimates than did the standard task (note also that *F* values

TABLE 2

MEAN OF LOG MEAN TIME ESTIMATE OVER ALL TRIALS FOR EACH GROUP

Standard Task	Comparison Task		
	Rest	Alphabet	Anagrams
Rest	1.42	1.54	1.73
Alphabet	1.25	1.43	1.62
Anagrams	1.27	1.41	1.46

were always larger for comparison task); as might be expected, the task *S* is working on at the time he must make his estimate is probably the more important. Perhaps it is also worth mentioning that these three same-task groups contributed to the heterogeneity of variance found in log scores over all trials; as compared to the other six groups, the variances of the three same-task groups were very small (and nearly identical).

At the end of the experiment *S*s were asked a number of questions; their answers yielded the following information. The *S*s estimated the length of the standard interval from as low as 2 to as high as 120 sec. Different *S*s thought the standard interval remained constant, progressively lengthened, progressively shortened, varied randomly. The *S*s were also asked what they thought about the task or tasks they had worked on. It was clear from *S*s' remarks that the great majority of them rather strongly felt that the rest task was unpleasant, boring, progressively more boring, and intellectually dull, while the anagram task was felt to be pleasant, interesting, and challenging. The alphabet task did not tend to elicit strong feelings one way or the other.

Discussion

The first purpose of the study was to test the prediction that times estimates, defined here as the time *S* worked on the comparison task, would vary inversely with the "level of behavior" of the standard task, but vary directly with the level of behavior of the comparison task. This prediction was confirmed, not only on Trial 1, but also on mean score over all trials, even though there was a significant interaction in this latter (all-trials) score. It is also worth noting that the difference between our extreme conditions (Rest-Anagrams vs. Anagrams-Rest) is in the same direction as that found by Hirsh, Bilger, and Deatherage (1956) when their *S*s were subjected to quiet or noise during standard and comparison intervals; quiet-to-noise produced much longer time estimates than noise-to-quiet. Furthermore, both studies found that this "order-of-presentation" effect was reduced when less extreme points along the dimension (level of behavior in the present study, decibel level of the noise in the Hirsh study) were compared.

The second purpose of the study was to check a finding of Kleiser (1953), who reported that time estimates became progressively longer over a series of trials. Although this finding was not confirmed, it should be noted that Kleiser varied only the standard task; his comparison task was always rest. Because of this, only two of the nine groups of the present study, Alphabet-Rest and Anagram-Rest, are comparable to Kleiser's groups. Inspection of the mean time scores of these two groups alone at the five trend points (blocks of three trials) did suggest that time estimates were becoming progressively longer over trials, although only slightly so. Therefore, these two groups were combined and the scores analyzed for trend. The analysis yielded an F for trend of 3.36, which is significant at the 5% level with 4 and 116 *df*. Thus, there was some evidence, although not strong, that time estimates tended to become longer when tasks at higher levels of behavior were repeatedly followed by rest. This finding does not seem to be a major exception to the over-all finding that time estimates remained quite constant over trials. Since appropriate control groups were not used, any progressive changes in time estimates over trials in either Kleiser's or the present study can only be said to be concomitant with, not due to, such variables as repeated experience with the procedure, improvement in task performance, etc.

It was suggested earlier that the comparison task exerted a more powerful influence on time estimates than the standard task. Further support for

this suggestion can be obtained by comparing the slopes of the curves (not shown) that appear if one plots the grand means for each task variable separately. For example, on Trial 1 (where the nonsignificant interaction permits one to combine groups on the other variable), the grand mean of log scores decreased from 1.54 with rest to 1.33 with anagrams, when these tasks were used during the standard interval. When the same tasks were used during the comparison interval, the grand mean increased from 1.26 with rest to 1.59 for anagrams. Thus, variation of level of behavior during the comparison interval seemed to have a somewhat greater effect than the same variation during the standard interval.

Finally, it is suggested that the three tasks used here do represent points along the poorly-defined dimension, level of behavior, at least insofar as they influence estimation of time. The tasks were originally selected because they seemed similar to the tasks used by Axel (1924) and Gulliksen (1927), i.e., rest similar to rest or tapping a pencil, reversed alphabet printing similar to number canceling or mirror reading, anagram solving similar to analogies or long division. Axel's or Gulliksen's tasks were not used because it was thought that the use of new tasks would, if the results were in the predicted direction, provide greater generality to the dimension of level of behavior. Since the major results were in the expected direction, the hypothesis that the present tasks do vary along the dimension receives some support.

Summary

Each *S* was given several trials in each of which he first worked on a standard task for a fixed interval of time, then worked on a comparison task until he estimated he had worked the same length of time as had elapsed during the fixed interval. Three tasks—rest, reversed alphabet printing, and anagram solving—were used in all combinations as both standard and comparison tasks. These tasks were assumed to represent, respectively, low, intermediate, and high degrees of a dimension called "level of behavior."

The results were: (*a*) As expected, time estimates varied directly with level of behavior of the comparison task, inversely with level of behavior of the standard task. These effects occurred over and above a significant interaction, found with one measure, between standard and comparison task variables. (*b*) In general, time estimates did not vary over trials, although there was some evidence that estimates became progressively longer when standard tasks at higher levels of behavior were repeatedly followed by rest as the comparison task.

References

Axel, R. Estimation time. *Arch. Psychol.*, 1924, No. 74.

Gilliland, A. R., Hofeld, J., & Eckstrand, G. Studies in time perception. *Psychol. Bull.*, 1946, **43,** 162–176.

Gulliksen, H. The influence of occupation upon the perception of time. *J. exp. Psychol.*, 1927, **10,** 52–59.

Harton, J. J. The influence of the difficulty of activity on the estimation of time. *J. exp. Psychol.*, 1938, **23,** 270–287.

Hirsh, I. J., Bilger, R. C., & Deatherage, B. H. The effect of auditory and visual background on apparent duration. *Amer. J. Psychol.*, **69,** 1956, 561–574.

Kleiser, J. R. The effect of habit formation, mental activity and gross muscle movements on the reproduction of time. *Dissertation Abstr.*, 1953, **13,** 886.

Rees, H. J., & Israel, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46,** (6), (Whole No. 210).

Weber, A. O. Estimation of time. *Psychol. Bull.*, 1933, **30,** 233–252.

Woodrow, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1224–1236.

(Received August 25, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

ONE DETERMINANT OF JUDGED SEMANTIC AND ASSOCIATIVE CONNECTION BETWEEN WORDS[1]

JOHN H. FLAVELL

University of Rochester

AND

ELEANOR R. FLAVELL

Rochester, New York

This past decade has seen a wealth of important investigations of word meaning and word association. As yet, however, very little is known about the variables of which strength of semantic and associative connection between words is a function. This study is concerned with one limited aspect of this problem, namely, the role of the logico-grammatical relation between words as a possible determinant or concomitant of judged semantic and associative connection. In a recent study, Flavell (1956) compared schizophrenics and normals as to judgments of semantic similarity among pairs of words bearing various logico-grammatical relationships to each other. Benedetti (1958) replicated this experiment with college student *Ss*, obtaining judgments of associative strength as well as meaning similarity. In neither investigation, however, was it possible to assess in any rigorous way the extent to which strength of semantic or associative connection was a function of the relation between two words, above and beyond the specific content meanings of each word in isolation.

The present study is a more systematic attempt to investigate the role of word-word relations, apart from word content, in three kinds of judgments of verbal connection: similarity of meaning, relatedness of meaning, and associative strength. Similarity and associative judgments were selected for study because of the obvious centrality of semantic and verbal habit connections in the area of verbal behavior; the inclusion of "relatedness of meaning" as a variable requires comment, however. In a recent paper, Bousfield, Whitmarsh, and Berkowitz (1958) have suggested that "word relatedness" is a more fundamental and inclusive bond between words than meaning similarity and should be a more potent determinant of mediational effects such as verbal transfer and clustering. Specifically, they argue that common antonyms, known to show strong transfer and clustering effects, ought to be considered as *related* but *dissimilar* in meaning. Thus, relatedness instructions were included in this study to find out whether, in naive *Ss*, they do, in fact, lead to different judgments from those given under similarity instructions.

Since the experimental procedure used was somewhat complex and detailed, a brief orienting summary follows. Eleven categories of relationships were used, each comprising a pair of words bearing the defined relationship to each other. Each pair could then be paired once with each other pair, thus making 55 pairs of pairs, called here a "set." A second set of 55 pairs could be constructed in the same way by using new word pairs to illustrate the same 11 categories.

[1] Part of this study was supported by a research grant (M-2632A) from the National Institute of Mental Health, Public Health Service. The authors wish to thank Alfred Bauer and Douglas Yeaple for their assistance in tabulating and analyzing data.

From the point of view of the logico-grammatical relationships involved, all sets constructed in this way are equivalent; from the standpoint of the particular word pairs used to exemplify these relationships, all sets are different. In this study, five distinct sets were used, thereby constituting a basic "test" consisting of 275 pairs of pairs. The *S*'s task, depending upon which group he was in, was to choose from each pair of pairs the one whose members seemed more similar in meaning, more closely related in meaning, or more closely associated. Since, within each set, each category was paired once with each other category, such choices in effect produce a ranking of categories (1 to 11) for each of the three judgments.

The rankings obtained in this way permit study of each type of judgment, and comparisons among judgments, with respect to the following: (*a*) extent of agreement among *S*s in judgments within each set; (*b*) more important, extent to which rankings are similar across the five sets—if they are similar for a given judgment, it can reasonably be concluded that level of judged connection is in part a function of the relationship rather than the separate words themselves; (*c*) the obtained rank hierarchies of relationships resulting from the various judgments.

Method

Subjects.—The *S*s were 222 sophomores enrolled in an introductory psychology course. They were rewarded with examination credit. Each of the three judgment groups consisted of 46 women and 28 men.

Test materials.—The 11 categories of logico-grammatical relationships are defined below. The five pairs of words used in this study to exemplify each relationship follow the definitions; the first pair listed after a given definition belongs to Set 1, the second to Set 2, etc.:

1. *Synonym* (SY). The members of each word pair have exactly or almost exactly the same referent (*big—large, steal—rob, car—automobile, couch—sofa, error—mistake*).

2. *Similarity* (SI). The members of each word pair are similar in meaning in the following way: the referents of both are closely aligned on some dimension, with the referent of the right-hand member occupying the more extreme position on this dimension (*small—tiny, smile—laugh, cool—cold, hungry—starving, damp—wet*).

3. *Supraordinate* (SU). The left-hand member of each word pair denotes a common class of which the referent of the right-hand member is a familiar class member, easily recognized as such (*bird—sparrow, animal—fox, furniture—table, fruit—apple, tree—oak*).

4. *Coordinate* (COO). The members of each word pair refer to familiar, easily recognized class members of a familiar class (*cow—horse, pipe—cigar, chair—table, knife—spoon, beets—peas*).

5. *Attribute* (AT). The right-hand member of each word pair refers to a quality or attribute generally recognized as always or nearly always characterizing the object denoted by the left-hand member (*lemon—sour, elephant—heavy, ice cube—cold, turtle—slow, mouse—small*).

6. *Contrast* (CON). The members of each word pair refer to opposite ends of a familiar continuum (*hard—easy, strong—weak, old—young, tall—short, loud—soft*).

7. *Action-of* (OF). The right-hand member of each word pair is an intransitive verb denoting a concrete action commonly associated with and frequently performed by the agent referred to by the left-hand member (*dog—bark, lion—roar, rabbit—hop, baby—cry, cat—purr*).

8. *Action-upon* (UP). The left-hand member of each word pair is a transitive verb denoting a concrete action commonly associated with and frequently performed upon the object referred to by the right-hand member (*sweep—floor, chew—gum, wash—hands, eat—bread, throw—ball*).

9. *Whole-part* (WP). The right-hand member of each word pair refers to a familiar object easily recognized as an important part of the familiar whole denoted by the left-hand member (*bird—wing, shoe—heel, hand—finger, tree—branch, bicycle—wheel*).

10. *Part-part* (PP). The members of each word pair refer to familiar objects easily recognized as important parts of a familiar whole (*pedal—handlebars, wall—floor, collar—sleeve, headlight—brake, arm—head*).

11. *Common use* (CU). The right-hand member of each word pair denotes an object commonly associated with and frequently acted upon by the agent referred to by the left-hand member (*dog—bone, farmer—tractor, baby—rattle, carpenter—nail, soldier—rifle*).

These particular 11 categories, culled from various sources, were selected for study because they were thought to be fairly unam-

biguous in definition, diverse in character, and frequently encountered in everyday language situations. They, of course, constitute only a small sample of the possible semantic relationships between words which could be abstracted from our language system. The word pairs used to represent each category were selected with the help of graduate student and staff judges in accordance with two guiding principles: (*a*) the words should be relatively common, familiar ones; (*b*) the five pairs for a given category should appear to display the defining characteristics of that category about equally well. Despite careful selection procedures, however, some difficulties remain; e.g., it is intrinsically more difficult to find homogenous pairs for some categories than for others, some word pairs are related in ways other than those specified by the category definitions. Problems of this kind are probably indigenous to any attempt to impose rigidly defined categories of semantic relationships upon a complex language system (Flavell, Draguns, Feinberg, & Budin, 1958).

Procedure.—The test taken by each *S* consisted of a page of detailed instructions (different for each judgment group) followed by 286 pairs of word pairs.[2] Of these 286, the first 11 were practice items; the remaining 275 were the critical ones for which choices were scored. The 275 items appeared in the test form in a random order which was the same for all *S*s.

The *S*s performing under similarity-of-meaning instructions were instructed to select from each two-pair item the pair whose members seemed to them "closer in meaning." For the word relatedness group, the instructions were essentially the same except that phrases like "more closely related in meaning," "more related to each other" were substituted for "closer in meaning," "A comes closer to meaning B than C does to D," etc. For the association group, the instructions were somewhat more detailed. A hypothetical continuous association situation was described in which an imaginary individual gives a sequence of associations to a single stimulus word. The instructions then specified that, for this individual, one would assume that the first associate given is most closely associated with the stimulus word, the second associate is next most closely associated, etc. The *S* was then directed to choose that pair of words from each pair of pairs which he thought would be the more closely associated for him, in the sense defined above.

Out of practical necessity, the test was self-administered on a take-home basis. The instruction sheet, in addition to specifying the task, also strongly exhorted *S* to take the test in one sitting that same evening, and without consultation with others.

Scoring.—Within each set, each *S* received a rank score (1 to 11) for each category based on the frequency (10 to 0) with which he chose this category in preference to the others with which it had been successively paired. All data analyses were based on these ranks.

Further experimentation.—In order to supplement data on similarity judgments, a second, minor experiment was conducted. Seven men and 19 women, mostly junior and senior psychology majors, were tested individually. These 26 *S*s were given the 11 category descriptions (somewhat less formally worded than the definitions presented above) plus five "examples" of each, consisting of the same word pairs used in the major study. In essence, the *S*s were instructed to rank the categories in terms of the generalized similarity of meaning they felt to be inherent in each category relationship. Care was taken to make the instructions as similar as possible to the similarity instructions in the major study. The data from this experiment made possible comparisons between similarity rankings derived from paired comparisons of concrete, individual word pairs, and rankings made directly on the basis of abstract definitions of relationships coupled with these same word pairs as concrete examples.

Results

The analyses described below were first performed on the data for each sex separately; since no meaningful pattern of statistically significant sex differences emerged, the analyses which follow involve total judgment groups only. Table 1 shows coefficients of concordance (*W*) for each judgment group (S = similarity, R = relatedness, A = association) expressing rank agreement among *S*s within each set separately and across the five mean

[2] All test material and instructions described in this paper have been deposited with the American Documentation Institute. Order Document No. 5 956, remitting $1.75 for 35-mm. microfilm or $2.50 for 6 × 8 in. photoprints. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress, Washington 25, D. C.

TABLE 1

KENDALL COEFFICIENTS OF CONCORDANCE (*W*) EXPRESSING AGREEMENT ON CATEGORY RANK ORDER WITHIN AND ACROSS SETS BY JUDGMENT

Judgment	Within Sets					Across Sets
	Set 1	Set 2	Set 3	Set 4	Set 5	
S	.53	.38	.58	.49	.55	.86
R	.44	.31	.48	.42	.38	.84
A	.30	.25	.15	.26	.25	.55

set rankings (Siegel, 1956). As the "within-sets" portion of the table shows, inter-*S* agreement is for each set highest for similarity judgments, next highest for relatedness judgments, and lowest for association judgments. Even the lowest *W*s in the table are highly reliable, however ($P < .001$). The across-set *W*s under S and R instructions are substantial and about equal; the corresponding *W*s for A judgments are definitely lower than the others. The group of 26 *S*s who made similarity rankings among categories directly from definitions and examples (S-Abst) showed an inter-*S* *W* (.58) comparable to the within-set correlations of the S group.

Table 2 shows the means of the individual set mean rank orders for each of the four judgment groups. Rank order correlations among the four rows of this table are as follows: S-Abst—S = .95; S-Abst—R = .93; S—R = .99; S-Abst—A = .40; S—A = .49; R—A = .57. It is clear that S-Abst, S, and R instructions produce highly similar category rankings, while A instructions produce a hierarchy decidedly different from any of these. Inspection of Table 2 suggests that relationships SY, SI, and SU show a relative drop in rank position under A instructions while OF, UP, and CON rise in rank.

In summary, similarity and relatedness instructions seem to have almost identical effects upon the judgment parameters studied here, with the exception of an apparent difference in inter-*S* agreement within sets. In both cases there is good evidence that a given relationship tends to maintain its rank status in the hierarchy despite variation in the words used to express the relationship. Furthermore, the two instructions produce highly similar relationship hierarchies. In the case of association instructions, on the other hand, inter-*S* agreement is lowest of all, rank status of relationships is less predictable from set to set, and the mean rank hierarchy of relationship differs substantially from that produced by the other two judgments.

DISCUSSION

Although the data presented in Tables 2 and 3 suggest that S and R instructions were, in all essentials, functionally equivalent in this study, it is not clear why

TABLE 2

MEAN ACROSS-SET RANK ORDER AMONG CATEGORIES BY JUDGMENT

Judgment	Category										
	SY	SI	AT	WP	SU	COO	OF	PP	UP	CU	CON
S-Abst	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0
S	1.0	2.1	4.9	4.0	5.0	6.0	5.8	9.4	8.6	8.2	11.0
R	1.0	2.4	4.3	3.6	6.2	7.1	5.0	10.2	7.8	7.6	10.8
A	4.3	5.2	3.8	3.6	8.6	7.8	2.9	10.4	4.6	8.2	6.6

this was so. It might have been expected, following the reasoning of Bousfield, Whitmarsh, and Berkowitz (1958), that *S*s would respond to instructions stressing meaning relatedness by promoting contrasts in the relationship hierarchy. It may, of course, be that changes in instruction wording, increasing the range and diversity of word relationships, etc., might well produce judgment differences not apparent in this study. On the other hand, it may simply be that the presence of the term "meaning," coupled with a general set towards comparing distances between words, tends to produce in naive *S*s a global strategy unaffected by connotative differences among expressions like "close or similar in meaning," "closely related in meaning," etc.

Granted that the two judgments did produce similar strategies in this study, strategies which showed a certain constancy across sets and *S*s, it is interesting to speculate as to what these might have been. One possibility is that most *S*s tended to adopt the strategy of judging words showing abstract, "similarity" relationships to be more related or similar than those expressing concrete, "contiguity" relationships (Flavell, 1956; Jakobson, 1955). Although the obtained hierarchy lends some support to this interpretation, there are some difficulties, e.g., WP and AT should be lower than SU in the hierarchy and this is not the case. Furthermore, such a strategy would not account for systematic rank differences *among* contiguity relationships or *among* similarity relationships. As one interesting example, subjects and their characteristic actions (OF) were almost always judged to be more similar and related than actions and their characteristic objects (UP).

As to the association judgment results, it seems reasonable to suppose that the lower level of within- and across-set agreement resulted from the fact that the instructions were more oriented towards idiosyncratic, immediate reactions in the individual *S* than was the case for R and S judgments. Although across-set comparisons were not possible in his study, Benedetti (1958) also showed that association judgments yield less inter-*S* agreement than meaning judgments. However, the relatively low correlations between A judgments and each of the other three (.40, .49, .57) are at variance with the results of an earlier study (Haagen, 1949). Haagen, using S and A instructions similar to those used here, found a correlation of .90 across adjective-adjective word pairs ranging from near-perfect to rough synonymity. This suggests that the relationship between semantic and associational judgments is a complex one which may, among other things, depend upon the form class of the words judged and the range of semantic distance sampled (Haagen's range is considerably narrower than that studied here). Finally, it has been shown that the hierarchy of relationships actually obtained under A instructions contrasts with the semantic hierarchies in several respects. That categories such as OF, UP, and CON should have higher rank status under A instructions appears reasonable, in view of the allegedly transitional, verbal-habit character of associational processes. Again, Benedetti (1958) also showed that association instructions favor verb-noun relations. What is surprising, however, is that coordinates and contrasts were not selected even more frequently than they were, in view of their very high frequency as responses in ordinary word association (Woodworth & Schlosberg, 1954). Evidence of this sort suggests considerable caution in interpreting judgments of associational strength as directly equivalent to operant word association as a means of assessing verbal habit connections.

According to one phrasing of the present findings, it has been shown that the semantic distance between two words is roughly predictable from prior knowledge of the logico-grammatical relationship uniting them. However, it may well be that these relationships do not really determine level of judged distance directly, but only through the operation of real determinants with which they are correlated or confounded. For ex-

ample, all pairs like *bird—wing, shoe—heel, tree—branch,* may be judged to be of approximately equal similarity or relatedness because the actual determinant or determinants of such judgments happen to operate at a more or less constant level of strength for all WP relationships. What might these determinants be? One possibility is that *S* responds in terms of some complex weighting of attributive or connotative similarity. Osgood, Suci, and Tannenbaum (1957, pp. 143–146) report a study by Rowan in which similarity judgments were successfully predicted from semantic differential D scores. Another possibility is that *S* makes his judgment in terms of a similarly complex weighting of implicit word association responses which the two judged words elicit in common. Cofer, in a recent experiment (1957), has lent some plausibility to this interpretation. Still another possibility, as yet hardly explored, is that an *S* judges two words to be similar in meaning to the extent that their referents tend to co-occur, "in nature," in the same spatio-temporal context. In a recent pilot study, the senior author presented in a fixed random order the 22 word pairs of Sets 1 and 2 to 15 graduate students under instructions to rate co-occurrence of referents on an 11-point scale. Each *S* first assumed an occurrence of the referent of A in an A-B pair and then rated the probability that B's referent would occur in the same context at the same time. After each word pair had been rated in this way, *S* made similar ratings of referent A's probability of co-occurrence, given referent B. The "co-occurrence score" for a pair was simply the sum of these two ratings. Inter-*S* agreement in this judgment task was comparable to inter-*S* agreement under S instructions: for Set 1, $W = .59$; for Set 2, $W = .46$. Rank order correlations between co-occurrence and *S* judgments were .78 for Set 1 ($P < .01$) and .56 for Set 2 ($P < .05$). If this correlation holds up over a more extended range of semantic distances, co-occurrence judgments may prove a fruitful third approach. Our present guess is that co-occurrence estimates, to the extent that they operate at all, probably operate multiplicatively with one or more other determinants in producing distance judgments.

Summary

The primary purpose of this investigation was to find out whether the logico-grammatical relationship obtaining between two words is a partial determinant of judged similarity in meaning, judged relatedness in meaning, and judged strength of associative connection. Five sets of 11 word pairs each were constructed, each pair in a given set showing one of 11 such relationships. Each of the 11 relationships was thus embodied in five different word pairs, one in each set. The *S*s made paired comparison judgments of similarity, relatedness, or associative connection on the 55 possible two-pair combinations of word pairs within each set, thus 275 judgments across all five sets. Intragroup agreement among *S*s within each set was highest in the case of similarity judgments, next highest for relatedness judgments, and lowest for association judgments. For both similarity and relatedness instructions there was a decided tendency for a given relationship to maintain its rank status across sets despite set to set variation in word content; rank stability was lower under association instructions. Similarity and relatedness instructions produced highly similar rankings of relationships; association judgments yielded a relationship hierarchy quite different from both. It was concluded that, for similarity and relatedness especially, judgments bearing on two words may be at least roughly predictable from prior knowledge of the logico-grammatical relationship involved.

References

Benedetti, D. T. The stratification of words in cognitive organization. *J. gen. Psychol.*, 1958, **58**, 249–258.

Bousfield, W. A., Whitmarsh, G. A., & Berkowitz, H. *Partial response identities in associative clustering.* Storrs: Dept. of Psychol., Univer. Connecticut, 1958. (Tech. Rep. No. 27, Contract Nonr 631–00.)

Cofer, C. N. Associative commonality and rated similarity of certain words from Haagen's list. *Psychol. Rep.*, 1957, **3**, 603–606.

FLAVELL, J. H. Abstract thinking and social behavior in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 208–211.

FLAVELL, J. H., DRAGUNS, J., FEINBERG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **57**, 1–7.

HAAGEN, C. H. Synonymity, vividness, familiarity and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

JAKOBSON, R. Aphasia as a linguistic problem. In H. Werner (Ed.), *On expressive language*. Worcester, Mass.: Clark Univer. Press, 1955. Pp. 69–81.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.

(Received August 26, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

THE EFFECTS OF VISUAL NOISE AND LOCUS OF PERTURBATION ON TRACKING PERFORMANCE [1]

WILLIAM C. HOWELL AND GEORGE E. BRIGGS

Ohio State University

In order to exercise vehicular control, *S* must be supplied with information regarding the course to be followed—the input signal (I) and the result of his own actions—the feedback signal (R). The pursuit type display provides I and R information separately by moving elements called the target and the cursor, respectively. The compensatory type display provides one fixed and one moving element, the difference between these two indicators representing error or the algebraic difference between I and R signals.

In a recent experiment, Poulton (1957) noted that whereas any degradation of information on a pursuit display resulted in a deterioration of tracking performance, the extent of performance decrement was a function of the locus of degradation. Thus, tracking was least proficient when the I information was subject to intermittent presentation, while performance was relatively less affected when the R information was degraded by intermittence. When the I and the R signals were simultaneously degraded, tracking performance approximated that found when the I information appeared intermittently on the display.

If Poulton's findings are truly indicative of the relative importance to tracking accuracy of the input and feedback information on a visual display, then a similar trend should result when other forms of display degradation are present. One source of degradation is *visual noise* which appears as random movement of the display element(s). This movement is uncorrelated with and independent of the input signal itself. It has been demonstrated that visual noise has a detrimental effect on tracking with a compensatory display (Briggs & Fitts, 1956; Briggs, Fitts, & Bahrick, 1957). When confronted with such a signal-in-noise condition, *S* is required to estimate and track the more coherent signal and disregard or "filter out" the random (noise) component of the displayed information.

The present investigation, then, was undertaken to test the generality of Poulton's findings for degradation other than intermittence, and for a compensatory as well as for a pursuit display. Visual noise was added to the elements of a pursuit tracking display in the following manner: (*a*) noise on the target or input element (I) only, (*b*) noise on the cursor element (R) only, and (*c*) noise on both I and R simultaneously. A fourth signal-in-noise condition was obtained by adding visual noise to the error signal on a compensatory tracking display (E). Further, the variables of input complexity and

[1] This research was conducted in the Laboratory of Aviation Psychology of Ohio State University and was supported in part by the United States Air Force under Contract No. AF 41(657)-70 with the Ohio State University Research Foundation, monitored by M. R. Rockway, who served as Scientific Officer for the Aero Medical Laboratory, Wright Air Development Center. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

amplitude of visual noise were included in the design to permit a more general interpretation of the results.

METHOD

Apparatus and scoring.—The *S* viewed a 5-in. cathode ray tube (CRT) display of tracking information, and he utilized a one-dimensional control device to generate his tracking response. A change in position of the control device produced an immediate and directly proportional change in the position of a cursor on the CRT. The control-display sensitivity was such that a 10° horizontal rotation of the control deflected the cursor 1 in. on the CRT. The input signals were displayed with a peak-to-peak amplitude of 4 in.; thus, *S* was required to rotate the control device through an arc of ±20° from its center position. This latter position was set at an angle of approximately 50° measured counterclockwise from the vertical plane of *S*'s thorax, and the axis of rotation was at *S*'s elbow. The control was not spring-centered and was subject to minimal friction. The tracking task was one dimensional and the CRT display contained two partially overlapping vertical lines, the target and the cursor elements, which moved only along the left-right axis. Two display modes were employed: a pursuit display and a compensatory display of tracking information.

Performance was measured by integrating the absolute values of tracking error during each trial. The error signal was scored prior to that point in the system at which the noise signals were introduced. Since tracking error appeared in the scoring circuits as a voltage, it was desirable to transform these raw error scores from arbitrary units to a meaningful metric expressed in scale units matching those of the tracking display, i.e., inches. Thus, the raw integrated error scores were divided by time, T = 60 sec., to provide a metric analogous to the average deviation of *S*'s error amplitude distribution, $\frac{1}{T}\int_0^T |e|dt$. This metric is termed average error. These average error scores were then transformed from arbitrary voltage units to a linear scale of inches of CRT movement. To illustrate, a score of .5 in. is interpreted as the average deviation of *S*'s error amplitude distribution if that distribution were drawn to the same scale as that of the visually displayed error.

Experimental conditions.—Three variables were manipulated systematically in this study: (*a*) input complexity (two levels), (*b*) magnitude of visual noise (three levels), and (*c*) locus of visual noise (four positions), a total of 24 combinations of experimental conditions. In addition, four noise-free conditions were included to provide a base line of tracking performance. Thus, *S* tracked both inputs with both the pursuit and compensatory display modes under noise-free display conditions. The two levels of input complexity included a single 10-cpm course and a more complex course consisting of a combination of 10 and 30 cpm sinusoids of equal magnitudes and zero phase angle. In either case, the peak-to-peak excursion of the input signal was ±2 in. on the CRT. The three levels of noise magnitude, N_1, N_2 and N_3, were described by random signals with normal amplitude distributions and average deviations of .03, .05, and .11 in., respectively.

The noise signals were produced by an electronic random signal generator (Clark, Fontaine, & Warren, 1953). The signals were peaked at 30 cpm with a bandwidth of 15–60 cpm. Attenuation at the bandwidth was 24 db per octave. The above estimates of average deviation were based on 15-min. records of the noise signals. Of the four loci of visual noise, three appeared on the pursuit display so as to show perturbation of (*a*) the target element only (I), (*b*) the cursor element only (R), and (*c*) both I and R elements simultaneously. As the fourth locus of perturbation, visual noise was added to the tracking error signal (E) on the compensatory display. When visual noise appeared simultaneously on the I and R elements, the noise signal added to the I element was statistically equivalent to but independent of that added to *S*'s output (R).

Subjects and procedure.—Twenty-four male undergraduate volunteers were paid for service in this experiment. Each *S* participated in one practice and four experimental sessions, and over the latter, each *S* tracked under all 24 combinations of noise amplitude, noise locus, and course complexity. The initial portion of the practice session consisted of 16 30-sec. tracking trials under N_0 conditions. This enabled *S* to approach asymptotic levels of proficiency under noise-free conditions for both pursuit and compensatory display modes with both the simple and the complex inputs. These trials were followed by 16 65-sec. trials also under N_0 conditions. There were four-trial blocks for each of the four combinations of display mode and input complexity. The performance measured during these 16 trials established the base line against which to compare proficiency under the N_1, N_2, and

N_3 noise conditions (see N_6 in Fig. 1 and 2). The scheduling of an additional set of 16 N_0 trials in this first session was further assurance that *S* had attained near-asymptotic skill at the basic tracking task.

Each of the remaining four sessions involved a total of 24 65-sec. trials. Locus of visual noise defined these four sessions (I, R, I and R, and E), and within each session *S* tracked four consecutive trials on each of the six combinations of input complexity and noise amplitude. An intertrial rest interval of 25 sec. was used with a 2-min. rest between blocks of four trials. Order of the four noise loci was counterbalanced across the last four experimental sessions via a latin square design, while order of conditions within a session was determined from a table of random numbers. The use of 65-sec. trials permitted a 5-sec. warm-up period prior to the scoring of tracking performance.

Instructions to *S* emphasized that he was to disregard the random variations in position of the display element(s) and track the more coherent signals. During the experimental sessions *S* was given explicit instructions concerning the particular condition to follow in each block of trials, and he was reminded frequently to "track through the noise." Knowledge of results in the form of integrated absolute error was given verbally immediately following each trial, and during the intertrial rest intervals *S* was shown (but not permitted to track) the noise-free display and the noise-free input to be tracked on the subsequent trials.

Results

The major results of this study are summarized in Fig. 1 and 2, where tracking proficiency is shown as a function of the magnitude of visual noise with locus of perturbation as the parameter. The data summarized in Fig. 1 were taken with the single 10-cpm input, while Fig. 2 represents tracking the more complex 10 + 30 cpm course. Each point in these figures is the sum of four 60-sec.

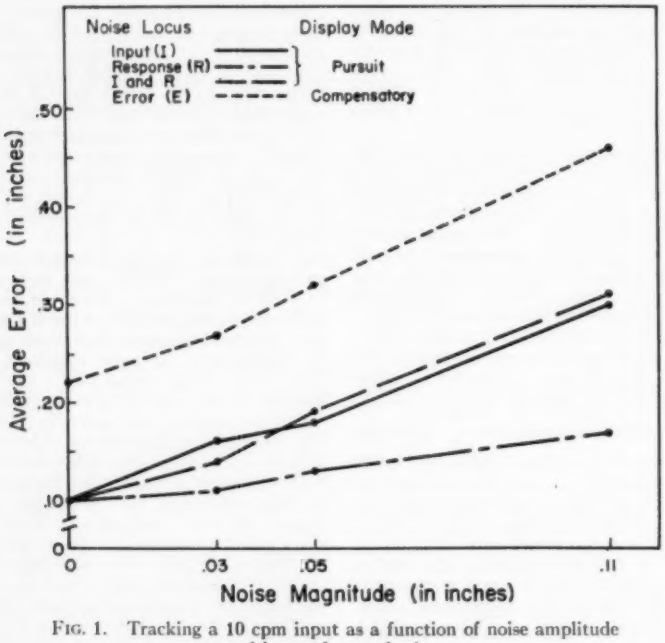

FIG. 1. Tracking a 10 cpm input as a function of noise amplitude and locus of perturbation.

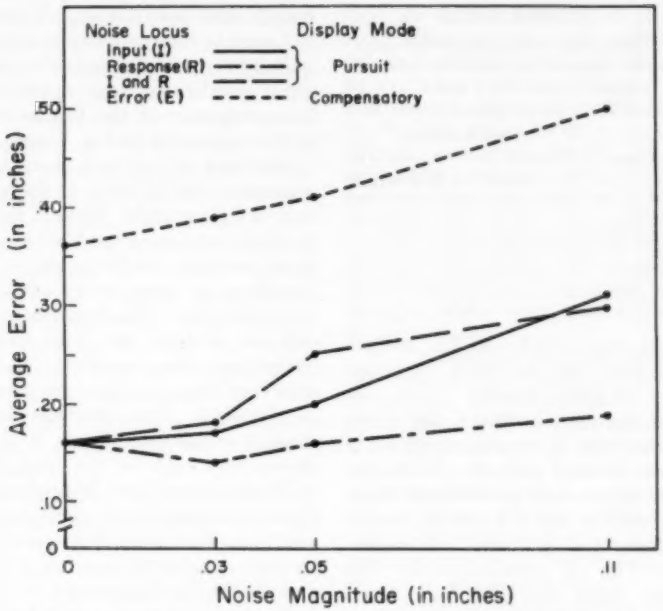

FIG. 2. Tracking a 10 + 30 cpm input as a function of noise magnitude and locus of perturbation.

trials for each *S*, averaged over the 24 *S*s. The average deviation metric discussed above serves to describe performance here.

From visual inspection of Fig. 1, it is apparent that, as expected (Briggs & Fitts, 1956; Briggs et al., 1957), an increase in the magnitude of visual noise resulted in decrements in tracking proficiency. Further, the extent of such decrement is dependent on the locus of perturbation: adding noise signals to the cursor element (R) produced the least decrement in tracking proficiency, while noise on the target element (I), on the I and R elements simultaneously, or on the error signal (E) resulted in a far greater reduction in proficiency. The results obtained with the complex input are comparable in this respect (see Fig. 2).

In order to illustrate this differential effect of visual noise, a linear equation was fitted to each of the eight functions shown in Fig. 1 and 2. Such a function provides a reasonable fit over the range of noise magnitudes studied here. However, data reported elsewhere (Briggs & Fitts, 1956) indicate that the initial portion of these functions is positively accelerated. Thus, the use of a linear equation here is justified only as an expedient to demonstrate differences in slope of the several noise locus functions, and does not suggest the general relationship between tracking error and noise magnitude. Table 1 presents the slope constants for these curves as determined by least squares fits. It is apparent from the table that for the simple input, the I and the I and R curves have nearly

TABLE 1

SLOPE CONSTANTS FROM A LEAST SQUARES FIT OF LINEAR EQUATIONS TO THE FUNCTIONS OF FIG. 1 AND 2

Noise Locus	Input Complexity	
	10 cpm	10 + 30 cpm
Input (I)	1.8	1.4
Response (R)	0.6	0.3
I & R	2.1	1.3
Error (E)	2.2	1.3

identical slopes (1.8 and 2.1, respectively), while that for the E condition is only slightly greater (2.2). In marked contrast, however, the curve describing the R condition shows a slope constant of only .6. With the complex input, nearly identical slopes were found for the I, I and R, and E conditions: 1.4, 1.3, and 1.3 respectively. The R condition for the complex input has the least slope (0.3).

Finally, it is interesting to note that the slope constants for the simple input are uniformly higher than those for the complex input. This results, of course, in a convergence of the functions, for comparable noise locus conditions, at the higher magnitudes of visual noise. This result was not predicted, since, within the range of noise amplitudes sampled, the increased complexity of the 10 + 30 cpm input was expected to make it more difficult for *S* to estimate this signal when obscured by noise as compared with the simple 10 cpm course. An a posteriori explanation for this finding is based on the fact that if *S* did not move his control device at all during a tracking trial, his error score would be lower for the more complex input. As calibrated, the score for the complex input would have been .7 in., while the average error metric with the simple input would have been 1.1 in. Thus, even if *S* were to filter out the noise equally well in the simple and the complex input conditions, one would expect a convergence of the functions such as that apparent in Fig. 1 and 2.

Analyses of variance were applied separately to the data of the pursuit and compensatory tasks. Separate analyses were required here as a combined analysis would result in a confounding of noise locus effects and task difficulty. Further, the analyses did not include the data from the initial noise-free condition, N_0. The results of these analyses are presented in Table 2. The analysis for the pursuit conditions (I, R, I and R) shows that each of the main effects of Noise Locus, Noise Magnitude, and Course Complexity, is significant. In addition, the Locus × Magnitude interaction is significant at $P < .01$, and the Course Complexity × Magnitude interaction is significant at $P < .05$.

The N_0 conditions were not included in the analyses of variance reported in Table 2 because, as indicated above, N_0 was experienced

TABLE 2

ANALYSES OF VARIANCE OF THE DATA DEFINING FIG. 1 AND 2

Source	*df*	*F*	
		Pursuit	Compensatory
Noise magnitude (M)	2	317.72**	118.54**
M × *Ss*	46	(2.48)[a]	(4.03)[a]
Noise locus (L)	2	18.25**	
L × *Ss*	46	(21.02)[a]	
Input complexity (C)	1	9.82**	61.66**
C × *Ss*	23	(6.50)[a]	(6.48)[a]
M × L	4	26.08**	
M × L × *Ss*	92	(2.13)[a]	
M × C	2	4.66*	4.82*
M × C × *Ss*	46	(2.16)[a]	(6.98)[a]
L × C	2	0.85	
L × C × *Ss*	46	(4.63)[a]	
M × C × L	4	2.28	
Residual	92	(1.72)[a]	

* $P < .05$.
** $P < .01$.
[a] Error term.

prior to N_1, N_2, and N_3, and therefore N_0 did not appear in the systematic combination of noise level, noise locus, and input complexity over Sessions 2–5. Instead, separate t tests were made between the N_0 and N_1 conditions for each of the noise loci. The results of these tests for the simple input (Fig. 1) indicate that all differences between N_0 and N_1 were significant ($P < .01$) with the exception of the R condition ($P > .05$). For the complex input (Fig. 2), however, no significant differences were demonstrated between N_0 and N_1. Additional t tests were then applied to the $N_0 - N_2$ differences. In this instance, all differences were significant ($P < .01$) except that for the R condition.

Subsequent to the above analyses, the Duncan Range Test (1955) was applied to the data defining the several functions in Fig. 1 and 2. Again, the N_0 data were excluded. As a result of this application and from the above t tests, it was established that the I, I and R, and E functions have significant slopes ($P < .05$), whereas none of the points on the R functions differ from the noise-free condition ($P > .05$). Thus, the two slope constants for the R condition, as listed in Table 1, do not differ significantly from zero.

Discussion

The above results clearly support the hypothesis that the various display elements contribute unequally to the accuracy of controlling responses of the human operator. As predicted from Poulton's findings (1957), performance suffers markedly when the target information alone (I) or the target and cursor information (I and R) is obscured simultaneously by visual noise, whereas a much smaller performance decrement occurs when the same perturbation is added to the feedback information alone (R). In addition, the prediction that noise added to the error information (E) in a compensatory display would have an effect similar to that for I and R in a pursuit display is upheld. Unlike the Poulton study, however, no reliable difference was found between the I and the I and R condition, and the decrement resulting from the addition of noise to the R element was not significant. Thus, it appears that degradation of the visual feedback information (R) by intermittence has a more detrimental relative effect on performance than degradation by noise.

The implication of these findings is that visually coded feedback of response information is less critical to operator performance than is input information. Since both the present study and that reported by Poulton lead to the same general conclusion, it is suggested that this differential effect is a general principle applying to a variety of conditions which result in a degradation of the display of information signals.

This unequal effect of display degradation may be explained, in part, by considering the alternatives available to S under those conditions when visual noise is added to the input signal, as compared to the alternatives when it is added to the feedback element. In the former, S may either filter out the noise as instructed, i.e., he attempts to discriminate signal from noise, or he may track the noise signal itself to a certain extent, in which case tracking error would be directly proportional to the magnitude of the noise. However, when the feedback information is obscured, another alternative is available: S may partially disregard the visually coded feedback information and rely to a greater extent on proprioceptive cues. That it is possible for him to learn to utilize these cues is suggested by Poulton (1957).

The use of proprioceptive cues as a source of response feedback information is particularly appropriate in this tracking task: S enjoyed positional control over the system, and thus the amount of movement of the R element on the

pursuit display was directly proportional to the extent of control device rotation. As such, there was maximum agreement between the visually coded feedback information and that available from the amplitude cues which arise from arm movement. Birmingham and Taylor identify this as a situation with a high level of "stimulus-response integrity" (1954). It has been demonstrated elsewhere (Bahrick, 1957) that practiced *S*s can effectively utilize such proprioceptive cues in direct positioning tasks. It is therefore logical to assume that the present *S*s were able to rely on these cues and thereby maintain proficient levels of performance, even with large magnitudes of visual noise on the R element of the pursuit display.

If this hypothesis is correct, the superior performance under the R condition over that with the other locus conditions would be expected to decrease as the "integrity" between visual and proprioceptive feedback information is reduced. Thus, if *S* were required to track through unquickened velocity or acceleration dynamics (Birmingham & Taylor, 1954), rather than with a positional control as in the present study, then proprioceptive cues arising from control movements would not be directly proportional to displayed feedback information. As such, *S* might find it difficult to substitute proprioceptive for visual cues, and the effects of visual noise on the R condition would be expected to approach that on the I and E conditions. Additional research is required to test this hypothesis.

It has been observed elsewhere that, in the range of input frequencies utilized here, compensatory modes of information display result in less proficient tracking than the pursuit mode (Chernikoff & Taylor, 1957). The same result is apparent in the present data. As indicated earlier, however, whereas the intercept values of the E and the I and R functions in Fig. 1 and 2 are significantly different, the slope constants are approximately the same (see Table 1). Thus, the addition of visual noise to the E element produces performance decrements of the same order as those obtained when either the I or the I and R elements are perturbed. This finding conforms to the present hypothesis in that *S* cannot substitute proprioceptive feedback cues for visually coded error information (the compensatory mode) any more than he can do so for visually coded input information (the pursuit mode.)

It may be concluded that *S* performs very inefficiently when required to track a coherent path through noise. When unable to substitute proprioceptive cues, as in the I, E, and the I and R conditions, it is apparent that *S* actually tracks the noise signals to a considerable extent (Fig. 1 and 2). As a demonstration of this, consider the N_1 point on the I function in Fig. 1. The average error at that point is .15 in., and this actually exceeds an average error which would be predicted by assuming that *S* attempts and succeeds in tracking perfectly the N_1 signal (.10 + .03 = .13 in.). An examination of the other points in Fig. 1 and 2 reveals that average tracking error in most cases approximates and even exceeds the error level predicted, as in the above example, from the average deviation measures of the noise magnitudes. Thus, *S* was not able to disregard the random perturbation of the input display elements despite near-asymptotic levels of skill on the basic tracking task.

Summary

Twenty-four *S*s each tracked four consecutive 65-sec. trials under 28 combinations of visual noise magnitude (four levels including a noise-free condition), locus of perturbation (four loci), and input complexity (two levels). Visual noise was defined as random perturbation of elements of the tracking display, and included noise on (*a*) the target element alone (I), (*b*) the cursor element alone (R), (*c*) both the I and R elements simultaneously, and (*d*) the error indicating element (E) of a compensatory display. A pursuit display mode was employed for the I, R, and I and R noise loci. The two inputs included a simple 10 cpm course and a more complex 10 + 30 cpm target course. The noise signals had a peak frequency of 30 cpm with a two octave band-

width. Integrated absolute tracking error served as the performance metric.

The results show (*a*) error increased in proportion to the magnitude of visual noise, with the R condition having the least decrement and the I, I and R, and E conditions having approximately the same decrement, which was significantly greater than that with the R condition, and (*b*) the over-all levels of tracking error were lower for the simple input compared to those with the more complex course; however, a convergence of the functions appeared at the higher levels of noise magnitude. It is concluded that *S* did not achieve a proficient filtering level and probably tracked the visual noise to a considerable extent. Finally, the agreement between these data and the results of Poulton on intermittence of comparable display elements (1957) suggests a general principle that degradation of visually displayed information in a continuous control task will have a differential effect on performance depending on the locus of such degradation.

REFERENCES

BAHRICK, H. P. An analysis of stimulus variables influencing the proprioceptive control of movements. *Psychol. Rev.*, 1957, **64**, 324–328.

BIRMINGHAM, H. P., & TAYLOR, F. V. A human engineering approach to the design of man-operated continuous control systems. *U. S. Naval Res. Lab. Rep.*, 1954, No. 4333.

BRIGGS, G. E., & FITTS, P. M. Tracking proficiency as a function of visual noise in the feedback loop of a simulated radar fire control system. *USAF, Pers. Train. Res. Cent. Res. Rep.*, 1956, No. AFPTRC-TN-56-134. (ASTIA No. 098911.)

BRIGGS, G. E., FITTS, P. M., & BAHRICK, H. P. Learning and performance in a complex tracking task as a function of visual noise. *J. exp. Psychol.*, 1957, **53**, 379–387.

CHERNIKOFF, R., & TAYLOR, F. V. Effects of course frequency and aided time constants on pursuit and compensatory tracking. *J. exp. Psychol.*, 1957, **53**, 285–292.

CLARK, J. R., FONTAINE, A. B., & WARREN, C. E. The generation of a continuous random signal for use in human tracking studies. *USAF, HRRC Res. Bull.*, 1953, No. 53-40.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

POULTON, E. C. On the stimulus and response in pursuit tracking. *J. exp. Psychol.*, 1957, **53**, 189–194.

(Received August 27, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

TWO-CHOICE BEHAVIOR OF RHESUS MONKEYS IN A NONCONTINGENT SITUATION [1]

WILLIAM A. WILSON, JR.

University of Colorado

AND

A. ROBERT ROLLIN

University of Connecticut

Several experiments (Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951; Humphreys, 1939; Jarvik, 1951) on human behavior in probability learning situations have produced results compatible with the so-called event-matching proposition, that the asymptotic distribution of responses in a two-choice situation will mirror the distribution of reward. In other studies (Bush & Mosteller, 1955, pp. 294–296; Detambel, 1955), conflicting results have been obtained: *S*s have chosen with a probability approaching 1 that response which is rewarded on the higher percentage of the trials.

Attempts have been made to explain these different results in terms of different experimental procedures, using the distinction between contingent and noncontingent situations. In the contingent situation the information available to *S* after he makes a response depends upon which response was made; specifically, *S* can learn only whether or not the response he made on a given trial was rewarded on that trial. In the noncontingent situation, the information available to *S* after he makes a response does *not* depend upon which response was made; he can see whether or not the other response would have been rewarded if it had been chosen. With some success it has been shown (Bush & Mosteller, 1955) that noncontingent experimental procedures were employed in human experiments leading to event-matching behavior and contingent procedures in those producing 100% selection of the more favorable alternative.

Most of the animal studies (Brunswik, 1939; Stanley, 1950) previously reported have used a contingent procedure. In a recent experiment, Bush and Wilson (1956) designed a noncontingent procedure for paradise fish. The experimental *S*s could see *E* place food in the correct goal compartment after each trial regardless of the choice *S* had made. These *E*s compared two specific models of the Bush-Mosteller stochastic learning model. According to an "information model," each trial increases the tendency of *S* to choose on future trials the alternative rewarded on that trial, but the influence of the placement of reward on a given trial is independent of *S*'s choice on that trial; this model predicts that *S*s will display event-matching behavior at asymptote. According to a "secondary reinforcement model," the response made on a given trial is reinforced whether or not *S* consumes the reward—by primary reinforcement if *S* has made the "correct" response, and by (less effective) secondary rein-

[1] This research was done at the Institute of Living and was supported in part by the Research and Development Division, Office of the Surgeon General, Department of the Army, under Contract No. DA-49-007-MD-763.

forcement if *S* has made the incorrect response and can only see the reward. This model predicts that each *S* will, at asymptote, choose the more favorable alternative with a probability of either 1 or 0, and that a higher proportion of *Ss* will have an asymptote of 1. Bush and Wilson viewed their results as supporting the secondary reinforcement model.

The writers were interested in studying monkeys in a similar noncontingent experiment to determine whether this intermediate species would show behavior more similar to that of humans or to that of paradise fish. The behavior of *Ss* using the noncorrection procedure of Bush-Wilson led us to modify the procedure to allow for immediate self-correction.

Method

Subjects.—Ten immature naive rhesus monkeys served as *Ss* for the first part of the experiment. One died, and the experiment was completed with the remaining nine.

Apparatus.—The experiment was conducted in a Wisconsin General Test Apparatus (WGTA). When *S*'s door of the WGTA was raised, he faced a black horizontal surface upon which were two black plastic boxes separated by a transparent plexiglass barrier. The boxes, which served as covers for shallow food wells, were hinged at the back, weighted, and connected so that if the front of either box were slightly raised both would fly open. The arrangement was such that when *S* chose a box by raising it, he could see both food wells but could only reach immediately the one he had chosen.

Procedure.—During seven days of preliminary training *Ss* were adapted to the apparatus and learned to open the food well covers to obtain reward. All preliminary trials were forced, and equal experience with both boxes was provided. Following this period the various training series began. Each *S* was given 32 massed trials daily. Training continued without interruption until completion of the experiment. The *Ss* were approximately 22 hr. hungry and the reward was always a half-peanut.

Part 1.—The procedure was modelled after that of Bush and Wilson (1956). Subjects

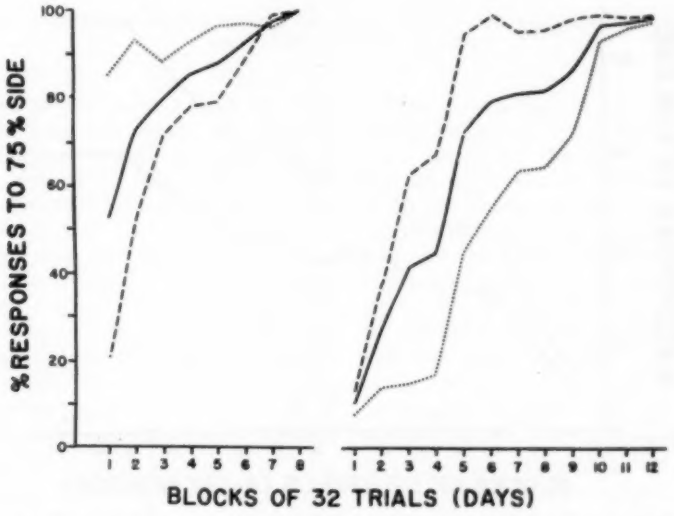

FIG. 1. Mean performance in noncorrection situation: (Left side: Part 1; Right side: Part 2). Dotted curves indicate five *Ss* originally trained to preferred side; dashed curves indicate remainder of *Ss*.

were presented with a position discrimination in which the more favorable side was rewarded 75% of the time, the less favorable side 25% of the time. The position of the reward was randomized within blocks of 32 trials, and the probabilities were "dependent," i.e., reward was placed on only one side in a single trial. Subjects were not allowed to correct their choices, but they could see both food wells after each choice. For half of the *Ss* the left was the more favorable side. Eight days of this training were given.

Part 2.—The procedure of Part 1 was repeated with the reward schedules reversed for each *S* so that the more favorable side on Part 1 was now less favorable, and vice versa. Training continued for 12 days.

Part 3.—The reward schedules of Part 2 were reversed, and a correction procedure was adopted. When *S* opened the unbaited box, he was permitted to reach around the barrier and obtain the peanut in the other food well. Training was continued for 60 days under these conditions.

Between Parts 3 and 4 each *S* was given an individual series of trials designed to change the probability of his choosing a given side from the value existing at the completion of Part 3 to approximately .50. On the trials of a single day, reward was either 100% on one side and none on the other, or randomly distributed 50% to each side. All *Ss* were required to meet a criterion of choosing each side at least 13 times in a daily series of 32 trials when the reward schedule was a random 50%—50%.

Part 4.—The training was identical with that of Part 3. For each *S* the side chosen more often on the later trials of Part 3 was made the less often baited (.25) side.

Part 5.—The side chosen more often on the later trials of Part 4 was made the less favorable one and the "noncorrection" procedure of Part 1 was used, for a period of 12 days.

Results

Parts 1 and 2.—All *Ss* showed an initial preference for the left side, and opened the left box on more than half of the trials on the first day. When they were incorrect, *Ss* usually

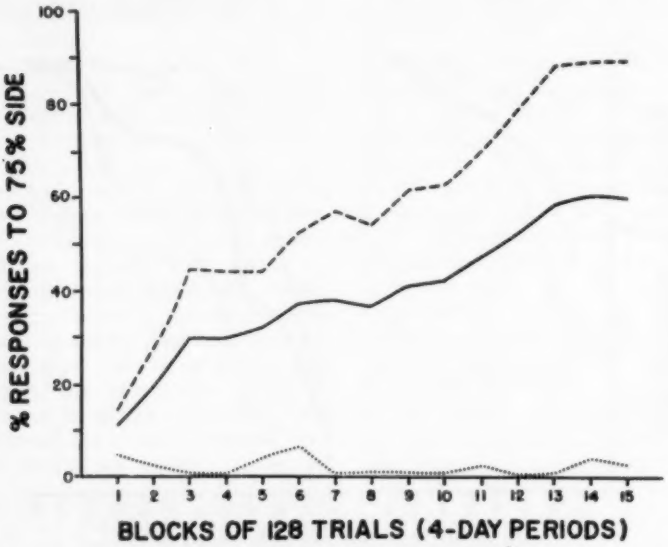

FIG. 2. Mean performance in correction situation, Part 3. Broken lines indicate subgroups defined on the basis of *Ss*' responses on the final block of trials.

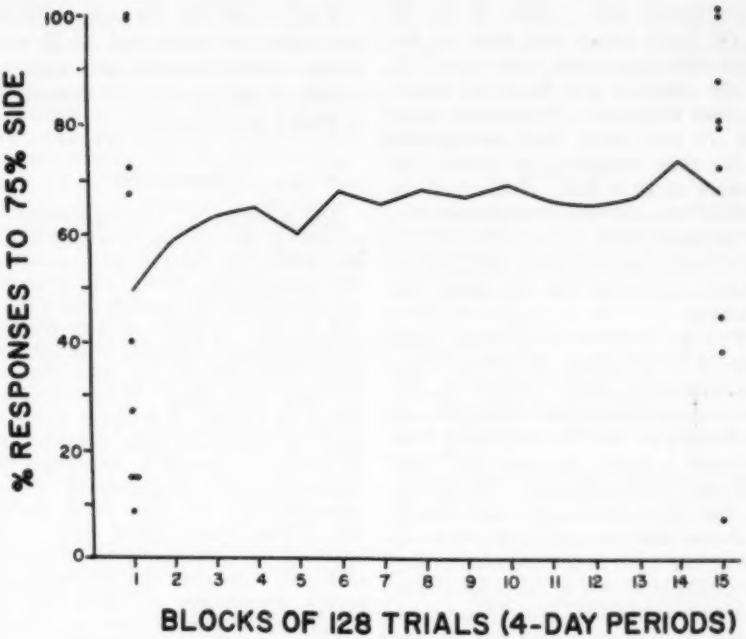

FIG. 3. Mean performance in correction situation, Part 4. Points indicate individual scores on first and last blocks of trials.

looked at the reward, and on early trials, tried to obtain it. Despite their preferences, all *S*s quickly began to go most of the time to the box that was rewarded 75% of the time; on the eighth day, all *S*s went to this more correct box on all 32 trials. The left side of Fig. 1 shows the daily means for the total group on this problem, and the means for the subgroups that had the left (top curve) or the right (bottom curve) box as the .75 alternative. As the right side of the figure shows, *S*s quickly reversed when the reward contingencies were reversed, and within 12 days chose the more favorable box almost all of the time.

Part 3.—In Fig. 2 the solid line shows the mean performance of the total group on the 1,920 trials of the first correction experiment. The mean proportion of choices of the .75 side reached approximately 60%, and had not obviously stabilized. The figure also shows the course of learning for two groups defined a posteriori on the basis of their performance on the last block of trials; the upper curve presents mean performance throughout this part of the experiment for six *S*s whose individual percentages of choice ranged from 74.2 to 100 on the last block; the lower curve is for three *S*s with final scores of 0.8%, 1.6% and 6.2% choices of the more

often baited side. None of the *Ss* in the latter group ever went to the more often baited side more than 25% of the time in any block of trials; in other words, any *S* that ever chose the .75 side more than one-quarter of the time continued to increase his choices of that side. Individual *Ss* who eventually went predominantly to the more often baited side initially responded approximately 100% of the time to the other side for from 3 to 46 days.

Part 4.—Individual *Ss* took from 5 to 54 days to meet the criterion for approximately equal choice of the two boxes. The mean curve (Fig. 3) for choices of the .75 side during Part 4 shows a slight increase and may still not be at asymptote. The points on the figure show the distribution of *Ss* on the first and last blocks of trials; the rank order correlation between these scores is .64. All of the *Ss* with between 20% and 80% choices of the more often baited side on the first block of trials increased their choices of this side.

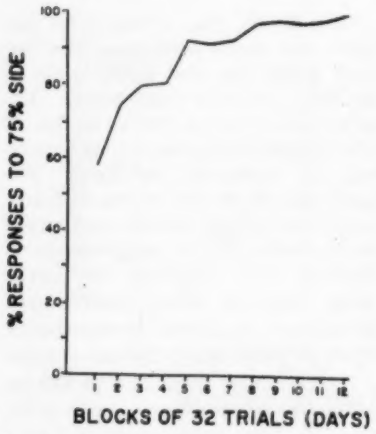

FIG. 4. Mean performance in noncorrection situation, Part 5.

Part 5.—When the noncorrection procedure was reinstated, all *Ss* went to the more favorable side approximately as quickly and consistently as in Part 1 (see Fig. 4).

DISCUSSION

The attempt in this experiment to produce a noncontingent situation was successful since by the construction of the apparatus and by the observed behavior of the *Ss*, on each trial *S* could and did see (*a*) what outcome resulted from his response, and (*b*) what outcome would have followed the other response if it had been made. Despite this attempt to provide a situation likely to result in event-matching behavior, all *Ss* in Part 1 adopted at asymptote the "most rational" procedure of always responding to the more often correct box. These results differ from those obtained in noncontingent situations with either paradise fish (where *Ss* approach asymptotes of either 0% or 100%) or humans (where an event-matching asymptote is reached).

The fact that all 10 *Ss* chose the more favorable side so regularly is inconsistent with the information model, which predicts that each *S* will choose this side 75% of the time. It is not so strong an argument against those models (the $\alpha_2 = 1$ model for contingent events or the secondary reinforcement model) which predict that *Ss* will go at asymptote to this side with probabilities of either 1 or 0, and with a larger percentage of *Ss* going with probability 1; for it may be that our sample was not sufficiently large to include any of that small percentage that would have gone consistently to the *less* favorable side. However, the fact that all *Ss* went to the .75 side when they began with *p*s of approximately 0 (Part 2) is further evidence against such models, since any *S* with a *p* of exactly 0 should not be affected by additional trials.

The possibility remains that *Ss* were responding just as they would have in a *contingent* situation where they were

given no information about the outcome of the response not made. One possible interpretation of such a suggestion is that visual information alone is not equivalent to the tactual and gustatory information available when *S* actually retrieves and eats a peanut, so that a noncorrection procedure for food reward cannot be a truly noncontingent one for the monkey. In an attempt to overcome this argument, the correction procedure of Part 3 was used. Once again there is no evidence that all *S*s are going to an asymptote of .75. On the other hand, there *are* *S*s that are responding to the more favorable side with *p*s approaching 0, but the course of behavior exhibited by the majority of *S*s suggests that these animals would have eventually reversed if training had been continued longer. Since all *S*s had begun Part 3 with *p*s approaching 0, it is not surprising, particularly considering the inefficiency of this training procedure, that some of them remained at approximately that level of performance throughout this correction situation.

In Part 4 we tried to provide a situation in which any *S*'s choice of the less correct side on a preponderance of the trials would be an impressive argument for a secondary reinforcement model, but were only partially successful in meeting the condition of having all *S*s begin with a p of 0.5. Granted the limitations of our test, there was *no* *S* who began this training with a p approximately .5 and ended it with a smaller value. The group curve is still rising at the end of the series, which suggests that individual *S*s are continuing to increase their proportions of choice of the .75 side. The asymptotic mean performance of the monkeys cannot be predicted with certainty, but once again it is possible that all *S*s would have eventually reached levels of approximately 100% choice of the more often baited side.

We cannot be certain that, in our *S*s, level of motivation, attitude, or other factors that have been said to affect behavior in these probability learning problems are comparable to those of humans in formally similar situations; we can say that one sustained attempt to get event-matching behavior from rhesus monkeys has not been able to produce it.

Summary

Ten monkeys were trained in a probability learning situation, one response being rewarded on 75% of the trials, and the alternative response on the remaining 25% of the trials. After each response, *S* also could see whether the response he did not make would have been rewarded if it had been made.

When correction of an unrewarded response was not permitted, all *S*s quickly reached an asymptote of approximately 100% choice of the .75 side, regardless of their initial probabilities of choice of that more favorable side.

When immediate correction was allowed, most *S*s slowly increased their proportion of choices of the .75 side. The data do not support the conclusion that all *S*s were approaching an asymptote of 75% choice of the .75 side; they are consistent with the suggestion that with sufficient testing all would have chosen the .75 side approximately 100% of the time.

The behavior of monkeys in this situation is not compatible with either an "information" or a "secondary reinforcement" model.

References

Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Bush, R. R., & Wilson, T. R. Two-choice behavior of paradise fish. *J. exp. Psychol.*, 1956, **51**, 315–322.

Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–104.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing

percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

STANLEY, J. C., JR. *The differential effects of partial and continuous reward upon the acquisition and elimination of a runway response in a two-choice situation.* Unpublished doctoral dissertation, Harvard Univer., 1950. Cited by R. R. Bush & F. Mosteller, *Stochastic models for learning.* New York: Wiley, 1955. Pp. 291–294.

(Received September 2, 1958)

Journal of Experimental Psychology
Vol. 58, No. 3, 1959

THE ROLE OF MUSCULAR TENSION IN THE RECALL OF INTERRUPTED TASKS

D. W. FORREST

Bedford College, University of London

Since the Zeigarnik (1927) effect was first discovered, the effects of interruption on recall have been investigated many times in attempts to refine the determining conditions. Two groups of these investigations are relevant to the present experiment, those attempting to find a possible physiological basis for the differential recall (Freeman, 1930; Smith, 1953), and those manipulating the attitude of *S*s to the interruption (e.g., Glixman, 1949; Lewis & Franklin, 1944; Rosenzweig, 1943).

It has been suggested that the state of the bodily musculature constitutes a possible physiological explanation of the Zeigarnik effect. Smith (1953) has shown that muscular tension in the active arm and in the speech muscles increased progressively up to the moment of completion or interruption of a simple mirror-drawing task. In the former case, tension then dropped sharply, whereas, when interruption occurred, it remained high for a short time. Thus it was possible to discriminate between an interrupted and a completed task at a muscular level, although in this experiment Smith made no attempt to investigate subsequent recall.

Smith hypothesized that this continuance of muscular activity after interruption reflects the self-maintaining action of a "phase sequence" (Hebb, 1949), which by reason of its incomplete organization is likely to dominate later recall. Finished tasks, on the other hand, will have correspondingly better-organized phase sequences which will have a greater probability of "short-circuiting," and so will be less likely to intrude into later remembering.

However, this formulation appears to ignore those experiments in which the attitude of *S* towards the task is experimentally determined. The results of such experiments (Glixman, 1949; Rosenzweig, 1943) generally show that when *S* is ego-oriented and interprets the interruption as failure, there is poorer recall of the interrupted *I* tasks, i.e., a reversal of the normal Zeigarnik effect where the *I* tasks are recalled more readily. If it can be shown that under these conditions also, an increase in muscular tension accompanies interruption, then it automatically follows that it is wrong to attribute any mnemonic significance to the higher tension. The present experiment was designed to check this possibility.

METHOD

Subjects.—The *S*s were 40 women undergraduates between the ages of 18 and 21 studying either sociology or mathematics, none of whom had heard of the Zeigarnik effect or suspected the purpose of the experiment. All *S*s were right-handed.

Apparatus.—Eight diagrams were used in a mirror-drawing task. The diagrams were the stenciled outlines of simple objects (a house, axe, car, spade, pencil, lamp, windmill, and ship) with additional details filled in with colored ink to ensure they could be easily recognized. The outlines measured 30 cm. in length in the case of the tasks which *S* was to be allowed to complete, and 35 cm. in those tasks which were to be interrupted. Interruption always occurred as near as possible to 5 cm. from the end of the task, so that approximately the same amount of work was performed in all cases.

The electromyograph was of private design (Forrest, 1956) and was used in conjunction with an integrator unit and a Kelvin-Hughes high-speed pen recorder which recorded on Teledeltos paper. A built-in calibrator was used to check the amplification before and after each set of tasks.

Bipolar electrodes of the sponge type soaked in saline were attached to the dorsal surface of the active arm half-way between the lateral epicondyle and the styloid process of the ulna (J. F. Davis, 1953). An earthing plate was taken from the upper arm.

Procedure.—All *S*s were told beforehand that they would be taking part in an investigation into muscular tension during drawing. They were assigned at random either to the ego- or task-oriented group. Instructions then differed for the two groups. The members of the task-oriented group were each told:

"You are a lucky subject as you're acting as a control for the other experimental group. We're not interested so much in how well you draw—in fact we're not interested in your ability at all—but rather in how good the *material* is for the task. We're going to record muscle potentials for this purpose. You may be stopped before you finish a task if we have got enough information about that task from what you've done."

Individuals in the ego-oriented group were read the following instructions:

"The purpose of this experiment is to measure your ability at a skilled task. Your performance will be compared with that of other members of the class and then we will show how it is associated with intelligence. If you should do very badly at a task you will be stopped before the end is reached. So interruption signifies failure. Is that clear?"

Both groups were then told:

"The task consists of a series of simple outline figures which have to be traced within a certain time limit while looking in a mirror. It is necessary for you to start and stop at the very instant that the command 'Start' or 'Stop' is given. Keep your pencil on the paper at all times and do not take it off when the command 'Stop' is given. Please keep it at the point where you stopped for a short period of time until you are asked to remove it. Are there any questions?"

All *S*s were reassured that there was no possibility of an electric shock. Then the electrodes were attached and a trial recording made while *S* practiced on a star pattern. The *S* was then given the first task of the series of eight. A rest period preceded the actual drawing. During this rest period, *S* was asked to hold her pencil at the beginning of the diagram and to plan out her future movements around the diagram. Continuous recordings were made from the beginning of the rest period until 5 sec. after interruption or completion. Interruption took place by *E* saying quietly, "Stop. Keep your pencil still." This was also said on those trials when *S* was allowed to complete the diagram, and in both cases a marker pen was activated on a second trace of the pen recorder at the moment of speaking.

Each *S* was interrupted on four of the tasks and allowed to complete the remaining four, both the order of presentation and the choice of tasks for interruption or completion being determined at random.

As soon as the drawing procedure was over, *S* was given a cancellation test at which she worked for the next 10 min. Most *S*s appeared to believe that muscular tension levels were still being recorded during this task.

The *S* was then asked to recall in any order she wished the objects she had traced. The time taken to recall each object was recorded and the order of recall noted. All *S*s reported that this came as a surprise, and none of them had tried to remember the objects when they were drawing them.

Quantification.—In view of Smith's (1953) finding that integrated action potentials over a period of 1 min. after interruption or completion were similar, but that averages over 2-sec. periods of time discriminated highly, it was decided to quantify in a similar way to his, i.e., by measuring spike deflections rather than integrator levels.

The number of spikes occurring in each period of 1 sec. after the moment of stopping were counted and 10 spikes chosen at random from each of these periods. These 10 spikes were measured to the nearest .5 mm. and these measurements averaged and converted into microvolts.

Results and Discussion

It is, first, essential to know whether the instructions were effective in bringing about a difference in the relative ease with which *I* tasks were recalled.

For each *S* the number of recalled *I* tasks minus the number of recalled *C* tasks was calculated, these were then ranked and a *U* test applied to

compare the two groups working under different conditions. A U value of 90 was obtained with $n_1 = n_2 = 20$, which gives a significance level of $<.01$. The task-oriented *Ss* showed the classical Zeigarnik effect and remembered more interrupted tasks, whereas the ego-oriented instructions were effective in causing a reversal of this result with the other group.

Now if the augmented tension found by Smith is the basis of improved recall, a difference in tension should be found between I and C tasks for the two groups. In the case of the task-oriented *Ss*, we should expect a higher average tension level to accompany the interruption rather than the completion of a task (as more members of this group recalled more interrupted than completed tasks), while in the ego-oriented group, the situation would have to be reversed.

Table 1 presents the mean tension levels for both groups over each second of the 5-sec. rest period that followed interruption or completion.

In the case of the task-oriented group, the interrupted tasks were accompanied by a higher tension level than the completed tasks over the first second of the rest period, and the probability of such a difference arising by chance is $<.05$, using a t test. This was to be expected, as it is in line with Smith's results. Thus, for this group, higher muscular tension is associated with better recall.

However, the ego-oriented group also showed an effect in the same direction and with a significance level of $<.01$. Yet this group remembered the completed tasks more easily, that is, those accompanied by less muscular tension.

It seems impossible to maintain the thesis that a high level of tension in the active arm is a necessary basis for improved recall. This is not to say that other loci of tension do not exist elsewhere in the musculature which may discriminate between remembered and forgotten tasks. But in the present experiment, higher tension levels appear to be the result of the physical act of interruption.

A comparison of the average muscular tension exerted by the two groups of *Ss* revealed no significant difference between them. Thus the fact that one group was given more formal instructions and was led to regard interruption as failure was not reflected in their muscular tension records.

When the course of muscular tension over the 5 sec. after the stop signal is examined, it can be seen that there is a considerable drop in tension over this period of time, and that the differences between I and C tasks are most marked for both groups at the beginning of the interval. Over the second, third, fourth, and fifth seconds the differences between I and C tasks are not significant for either group. It seems probable that if recording had been continued, this state of affairs would not have altered, as residual tension would have reached a true resting level. It is therefore unlikely that such a momentary phenomenon could have served as an adequate basis for a mnemonic process which was to last at least 10 min.

TABLE 1

MEAN TENSION LEVELS (μV.) OVER EACH SECOND OF REST PERIOD

Group	Task	Seconds After Stop Signal				
		1	2	3	4	5
Task-oriented	I	9.8	5.1	5.2	5.0	4.9
	C	7.0	4.9	4.7	4.1	4.2
Ego-oriented	I	11.2	6.4	4.5	4.7	4.4
	C	6.4	4.8	4.1	4.3	4.3

This theoretical difficulty is admitted by Smith, who likewise found a significant difference between *I* and *C* tasks in the 2-sec. periods immediately following the drawing, but not when longer postperiods were compared.

This is not a decisive drawback for any theory that utilizes the phase sequence type of explanation, as the neural activity presumably need not be reflected in peripheral tension change. Although the performance of a drawing task may be "organized" in the sense that tension progressively increases up to a maximum at the moment of stopping and then decreases over the next few seconds, such changes may be specific to this type of task; they are not necessarily found in nonmotor tasks (e.g., R. C. Davis, 1939; Forrest, 1956; Geldreich, 1953). There thus seems no good reason for supposing that the high muscular tension recorded after the interruption of a motor task is anything more than a motor phenomenon due to the sudden prevention of the continuance of a skilled movement.

Summary

This experiment was designed to discover whether the high level of muscular tension which has been shown to occur after the interruption of a motor task is a possible somatic basis for the Zeigarnik effect. To this end two groups of *S*s were employed, one task-oriented, the other ego-oriented. An EMG record was taken from the active arm during the performance of eight mirror-drawing tasks, half of which were interrupted.

As the ego-oriented group showed the expected reversal of the Zeigarnik effect, and as in both groups it was found that more muscular tension occurred after interruption than after completion, it was concluded that the increased tension was not a necessary concomitant of enhanced recall.

References

Davis, J. F. *Manual of surface electromyography.* Montreal: Lab. for Psychol. Stud., McGill Univer., 1953.

Davis, R. C. Patterns of muscular tension during 'mental work' and their constancy. *J. exp. Psychol.*, 1939, **24**, 451–465.

Forrest, D. W. A study of electromyographic changes associated with mental work. Unpublished doctoral dissertation, Univer. of London, 1956.

Freeman, G. L. Changes in tonus during completed and interrupted mental work. *J. gen. Psychol.*, 1930, **4**, 309–334.

Geldreich, E. W. Some physiological concomitants of mental work. *Psychol. Monogr.*, 1953, **67** (8), (Whole No. 358).

Glixman, A. F. Recall of completed and uncompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.

Lewis, H. B., & Franklin, M. An experimental study of the role of the ego in work. II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.

Rosenzweig, S. An experimental study of repression with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.

Smith, A. A. An electromyographic study of tension in interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 32–36.

Zeigarnik, B. Über das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85.

(Received September 15, 1958)

Journal of Experimental Psychology
Vol. 58, No. 2, 1959

REACTION TIME AS A FUNCTION OF FOREPERIOD DURATION AND VARIABILITY[1]

LAWRENCE KARLIN

New York University

In an early experiment, Woodrow (1914) studied the relationship between simple RT and foreperiod, i.e., the time intervening between warning signal and reaction stimulus. When foreperiods were constant for each block of trials (so-called "regular preparation"), he found that the 2-sec. foreperiod was optimal in that it yielded the shortest RT as compared to RTs obtained with foreperiods ranging from 1 to 24 sec.

Because subsequent experiments (Huston, Shakow, & Riggs, 1937; Klemmer, 1956; Rodnick & Shakow, 1940; Teichner, 1954) have not agreed in confirming Woodrow's results, the present study was concerned with a more detailed determination of the shape of the functional relationship between RT and foreperiod in the region of Woodrow's optimum.

A second purpose of the present study was to determine the way in which preparation to respond as indexed by RT develops during short foreperiods of varying length. It was hoped that the results would provide information concerning whether there were characteristic readiness curves such as those hypothesized by Woodworth (1938) for relatively long foreperiods.

METHOD

The present study consists of three experiments. Except where noted the procedure is the same for all three. Experiment II was performed to introduce some methodological refinements and to provide additional data to support the more detailed analysis required by the findings of Exp. I. Experiment III was performed as a further check on the discrepancy between the results of Exp. I and those obtained by Woodrow.

Subjects.—The *Ss* were 31 experimentally naive college students, eight each in Exp. I and II, and 15 in Exp. III.

Apparatus and stimulus conditions.—A 240-cycle tone of 33 db intensity (re .0002 microbar) and duration of .1 sec. was both the warning signal and reaction stimulus. In Exp. II only, the duration of this tone was shortened to .03 sec., and its intensity was increased to 49 db. The reason for the decreased duration was to minimize the possible ambiguity that may have occurred in reacting either to the onset or cessation of the longer tone of Exp. I and III. Intensity was increased because shortening the tone this much made it more difficult to hear. Three decade electronic timers were connected to provide the sequence, ABAC. Timer A controlled stimulus duration, Timer C, the rest period, constant at 7 secs., and Timer B regulated foreperiod duration. Each setting of Timer B was programmed by a stepping relay to provide "slightly" varying foreperiods when needed. The reaction movement was the upward release of a spring-loaded reaction key. RT was measured by a Springfield Timer calibrated in milliseconds. The *S* sat in one room, wearing a pair of headphones. In this room, which was adjacent to the room containing the apparatus, no equipment noises could be heard by *S*.

Procedure.—The following description applies only to Exp. I and II. The procedure of Exp. III will be given later along with a discussion of its results.

Each *S* in Exp. I and II attended 50-min. sessions on five consecutive days at about the same time each day. Only one experimental condition, C or V was used in each session. For four *Ss* the order was CVCVC, for the other four *Ss* it was VCVCV. In Cond. C, the foreperiods were .5, 1, 2, and 3.5 secs., and each one was repeated in two blocks of 21 trials each. There were thus 42 trials for each foreperiod, yielding a total of 168 RTs for the session and a grand total

[1] This research was supported by a grant from the New York University Arts and Science Research Fund.

of 840 RTs per *S*. Foreperiods were arranged in ascending-descending (AD) or descending-ascending (DA) order. The four *Ss* in each group described above were further subdivided into two groups who received either the DA or AD order. In Cond. V, the 21 trials in each block were divided equally among three foreperiods, with the median foreperiod corresponding to one of the foreperiods of Cond. C, while the other two foreperiods were 20% above and 20% below this median, respectively. The first trial of each block was always the median foreperiod; the remaining 20 trials consisted of the three foreperiods arranged so that each one followed itself and the other two about an equal number of times. The foreperiod durations were .4, .5, .6; .8, 1.0, 1.2; 1.6, 2.0, 2.4; 2.8, 3.5, 4.2 sec.

Condition V had two purposes: (*a*) to guard against attempts to synchronize reactions to foreperiod duration (Woodworth & Schlosberg, 1954) and (*b*) to provide a basis for the construction of the "readiness" curves previously referred to.

In both conditions, the first trial of each block was not used. After *S* was seated, he was given instructions which emphasized that he avoid premature reactions and have his two fingers (second and index) on the reaction key so that the contacts would be closed prior to the warning signal. When any premature reaction was made, *S* was so informed by an oscillating tone and the trial was repeated.

Results and Discussion

The first day's data were omitted from the results of Exp. I and II in order to equate the number of sessions per condition for each *S*, as well as to minimize practice effects.

The optimal foreperiod.—Table 1 summarizes the results of Exp. I and II on the relationship between RT and foreperiod. These results are also plotted in Fig. 1 where the RT-foreperiod function for Cond. C is shown by the broken line. The comparable function for Cond. V may be studied by considering RT for the median foreperiod in each of the four groups of foreperiods which have been connected by solid lines. The functions in both conditions show that RT rises from .5 to 3.5 sec. without inflection. Another way of describing the RT-foreperiod function in Cond. V is to average RTs for the three foreperiods in each set and plot the average RT as a function of median foreperiod. These averages are not plotted but are given in Table 1 where inspection shows that they follow a similar trend to that depicted in Fig. 1. Since no inflection was obtained, Woodrow's finding of a 2-sec. optimum is not confirmed.

There is also no evidence that the results of Cond. C were appreciably affected by prematurely originated reactions such as might have been expected if *S* were attempting to synchronize his reactions to foreperiod duration. Not only are the two functions similar in shape, but the actual number of premature reactions is very small and does not differ appreciably in either condition.

TABLE 1
RT as a Function of Foreperiod Duration and Variability. Experiments I and II

Cond.	Deviation from Median Foreperiods (%)	Duration of Median Foreperiod (Sec.)			
		.5	1	2	3.5
Experiment I (N = 8)					
V	−20	232	270	289	296
	0	233	256	273	274
	20	247	251	267	286
	Mean	237	259	276	285
C	—	221	256	271	280
Experiment II (N = 8)					
V	−20	217	247	272	263
	0	213	223	240	253
	20	224	225	237	244
	Mean	218	232	250	253
C	—	178	205	234	234

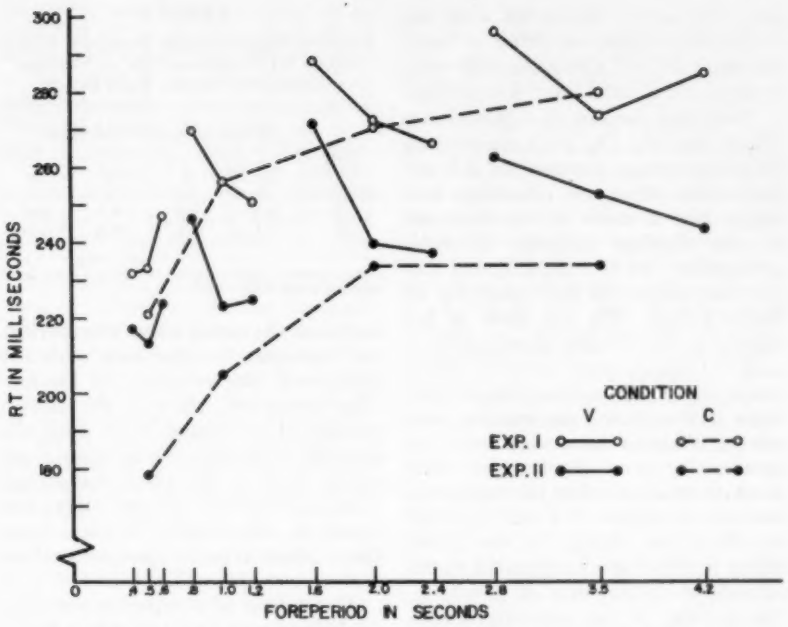

FIG. 1. Reaction time as a function of foreperiod duration in Exp. I and II. $N = 8$ in each experiment.

As a further check on the discrepancy between the present results and those of Woodrow, the possibility was investigated that "warm-up" effects in Woodrow's experiment produced the superiority of the 2-sec. over the 1-sec. foreperiod since he used only the ascending order of foreperiods in each session. Woodrow (1914, p. 54) checked the possibility of unbalanced practice and fatigue effects but did not feel that it was necessary for his purposes to counterbalance them. He seems not to have considered the specific possibility of warm-up effects. Accordingly, in Exp. III we tested 15 unpracticed *Ss* for only one session using the procedure of Cond. C except that the .5-sec. foreperiod was omitted and only the AD sequence was used.

As shown in Table 2, the results of Exp. I and II are confirmed, and no inflection in the ascending series which might be due to a "warm-up" effect is indicated. However, it is worth noting that six *Ss* did yield 2-sec. inflections in the ascending order as compared to one in the descending order. Since Woodrow only used three *Ss*, it may be that

TABLE 2

RT AS A FUNCTION OF FOREPERIOD DURATION: EXP. III ($N = 15$)

Series	Foreperiod Duration (Sec.)		
	1	2	3.5
Ascending	234	250	274
Descending	234	255	273

sampling factors interacted with his conditions to produce either a 2-sec. optimum or a nonsignificant difference between the 1- and 2-sec. foreperiods.

Foreperiod variability.—Woodrow (1914, pp. 28, 29, 35f.) maintained that the regular increase in RT for foreperiods of 2 sec. duration and longer was a result of the decrease in the absolute accuracy of time estimation. In fact the shape of this function suggested the operation of Weber's Law. On the basis of his argument, the foreperiod variabilities used in Cond. V were varied in direct proportion to the median foreperiods. Even if Woodrow's assumption concerning Weber's Law were only approximately true, it was felt that constant relative rather than constant absolute variability in Cond. V would introduce less change in the shape of this function when compared to the corresponding function of Cond. C. The validity of this rationale is supported by the fact that the functions for Cond. V and C in Fig. 1 show no tendency to diverge, even though the range of foreperiod variability in Cond. V increases from .2 to 1.4 sec.

Readiness and the effect of preceding foreperiods.—In setting up this study it was hoped that Cond. V would provide some information on how *S* prepares to react within each block of foreperiods since the slightly varying foreperiods within each such block might catch *S* at different stages of preparation. Thus the curves of Cond. V shown in Fig. 1, might be interpreted as empirical readiness curves (Woodworth, 1938) showing the way in which *S*'s readiness develops during a given block of foreperiods. Following this interpretation, the curves for the 1-, 2-, and 3.5-sec. blocks of foreperiods may mean that *S* is least ready (highest RT) when the reaction stimulus occurs at the end of a short foreperiod, but increases his readiness with the additional time provided by the median foreperiod. Beyond the median foreperiod, readiness, if it does not actually increase, is at least maintained above that of the short foreperiod. This description clearly does not apply to the curves of the .5-sec. block which in both experiments show a reverse tendency.

TABLE 3

F Ratios Derived from Variance Analyses of RT Variation Due to Varying Foreperiod Within Each Block

Exp.	Median Foreperiod Duration (Sec.)			
	5	1	2	3.5
I	2.1	5.2	9.9	8.9
II	1.1	4.4	14.0	1.6

$F_{.05}$ based on 2 and 14 *df* for greater and lesser mean squares, respectively, is 3.7.

The results of a separate variance analysis of each trend shown in Table 3 suggest that the trends for the three longer foreperiods are real, since five out of six *F*s are significant at much better than the .05 level. On the other hand, *F*s for the .5-sec. block are not significant at this level. In any case, we can conclude that the .5-sec. block of foreperiods yields results that differ from those obtained with the longer foreperiods, but we cannot say at this point that the variation within the .5-second block is systematic.

In interpreting the curves of Fig. 1 as readiness curves, we have neglected the fact that each of these curves is not based on conditions designed to promote preparation for a fixed foreperiod. In other words, the curves may not really describe the development of a uniform set or state of readiness because there is no single definable set characteristic of the three foreperiods that make up a given block. For example, both Klem-

mer (1956) and Woodrow (1914) have noted that RT is not only influenced by the length of a given foreperiod but also by the length of the immediately preceding foreperiod. Specifically, a short foreperiod preceded by a long one may catch *S* "napping," while a long foreperiod preceded by a short one will not have an adverse effect since *S*, influenced by the short foreperiod, gets ready more quickly and tends to maintain the readiness so developed.

It seemed possible that the curves for Cond. V in Fig. 1 could be explained partly in terms of these effects. For example, the consistently longer RTs obtained with the shortest foreperiods in the 1-, 2-, and 3.5-sec. blocks might be due, not to the length of foreperiod per se but to the fact that the shortest foreperiod would be preceded more often by longer foreperiods. On the other hand, this prediction could not be made for the .5-sec. block of foreperiods since the shortest foreperiod in this block did not yield the longest RT.

To explore the above suggestions, Fig. 2 plots readiness curves comparable to those in Fig. 1, except that it tries to make explicit the effects of the immediately preceding foreperiod. Each of these so-called readiness curves is based only on those foreperiods which have been immediately preceded by foreperiods of a given magnitude. For example, the L curves (solid line) show RTs only for those short, medium, and long foreperiods within each block that have been preceded by a long foreperiod.

The results shown in Fig. 2 support the suggestion that each of the curves for Cond. V in Fig. 1 are composites of different readiness curves. Except for the .5-sec. block of foreperiods, this interpretation is indicated by the uniform tendency of the curves for long preceding foreperiods to drop quite sharply in RT within each of the three longer blocks, as contrasted with the absence of any significant trend for those for short preceding foreperiods in the same blocks. The absence of any trend in the latter case may be due to the fact that the curve for short preceding foreperiods reaches a peak of readiness earlier than for long preceding foreperiods, so that readiness is not changing any more but is merely being maintained.

Examination of the curves obtained for the .5-sec. block of foreperiods suggests that the way in which preparation develops is also influenced by the effects of the preceding foreperiod. However, the trends differ radically from those obtained for the three longer foreperiods. Thus, the curves for long preceding foreperiods no longer display any significant trend and, in fact, show a slight tendency to vary in the opposite direction. In contrast, the curves for short preceding foreperiods are very significant statistically, showing a marked tendency to increase in RT beyond the median foreperiod. These results suggest that the patterns of preparation for this block of foreperiods may be different from those of the longer foreperiods in two respects: (*a*) a tendency to get ready as fast as possible which is superimposed on specific tendencies resulting from different lengths of preceding foreperiod, and (*b*) less ability to maintain a peak of readiness. The first conclusion is based on the fact that the curve for long foreperiods does not show an initially higher RT which characterized the curves of the three longer blocks of long foreperiods in both experiments. In effect, this means that in the .5-sec. block of foreperiods, *S* is not caught "napping" when a short period follows a long one. The second conclusion is based mainly on the sharp increase in RT displayed by the curve for short foreperiods. This result suggests that *S* is unable to maintain the peak of readiness shown by the curve for short foreperiods even for so short a time as an additional .1 sec.

While these interpretations are based on the results of both experiments, it is worth noting that the trends shown in Fig. 2 are much sharper in Exp. II.

Summary

The relationship between RT and foreperiod was studied under two conditions of foreperiod variability. Under one condition

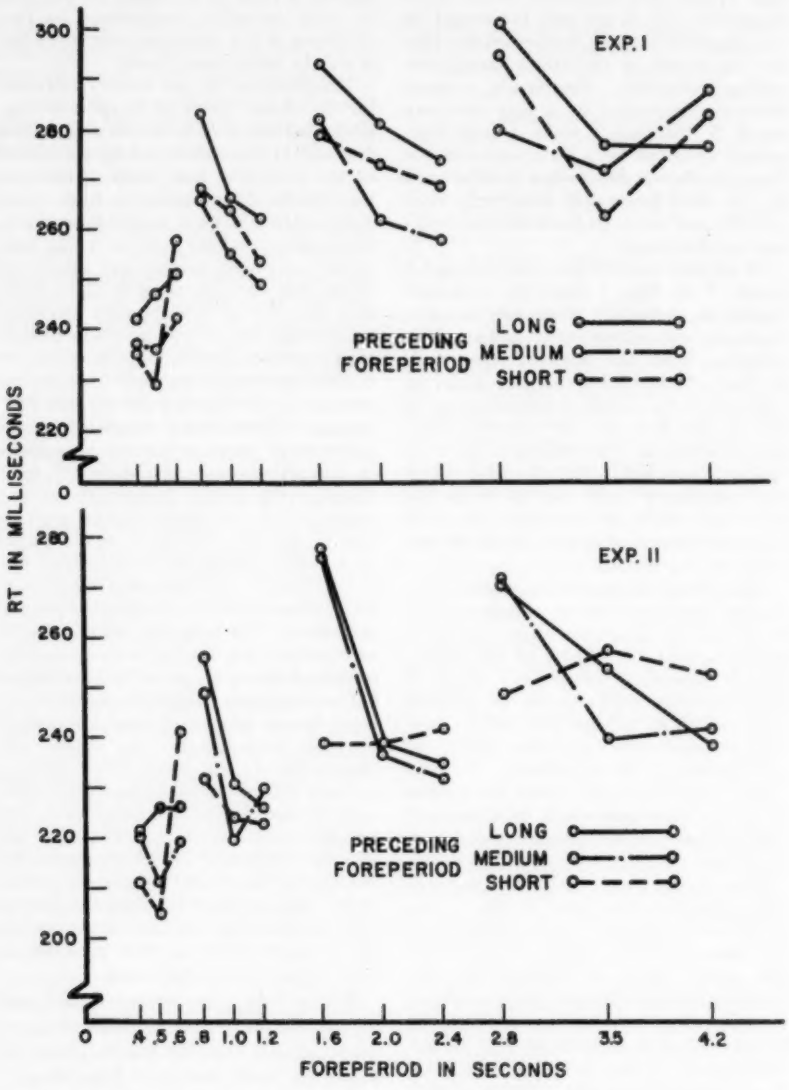

FIG. 2. Reaction time as a function of foreperiod duration in Exp. I and II showing the effects of various lengths of preceding foreperiod.

there was minimal objective variability, under the other there was 20% objective variability. The latter condition was used in an attempt to reduce the influence of prematurely originated reactions, and provide a method for studying the development of preparation. The results obtained suggested the following:

1. There is no evidence that prematurely originated reactions had an appreciable influence on either condition.

2. Woodrow's finding of a 2-sec. optimum was not confirmed.

3. The steady increase in RT as a function of foreperiod contradicts a previous conclusion (Teichner, 1954) that the optimum lies within a range of from 2 to 8 sec.

4. RT following a given foreperiod was influenced by the magnitude of the foreperiod immediately preceding the given foreperiod.

5. The development of a state of readiness followed a similar course during the three longer blocks of foreperiods which was different from the way in which it developed during the shortest block of foreperiods.

References

Huston, P. E., Shakow, D., & Riggs, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.

Klemmer, E. J. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

Rodnick, E. H., & Shakow, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214–255.

Teichner, W. H. Recent studies in simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

Woodrow, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17** (5), (Whole No. 76).

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology* (Rev. ed.), New York: Holt, 1954.

(Received September 15, 1958)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Journal of the German Psychological Association

Editors for Germany

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Editor for Austria
PROF. DR. H. ROHRACHER, Vienna

Editor for Switzerland
PROF. DR. R. MEILI, Bern

Redaction
DR. C. J. HOGREFE, Göttingen

Vol. 5, No. 3

ARTHUR L. BENTON. **The Benton Test**

WOLFGANG VON MOERS-MESSMER. **A Contribution on the Psychology of Form Gestalt**

EBERHARD ULICH. **Question of the Five-Day-Week in the Schools**

HORST MYERHOFF. **Experimental Research on the Psychology of Tubercular Diabetics**

KARL-HEINZ SEIFERT. **Researches on the Problem of Compensation in Manual Dexterity**

GOTTFRIED PETRI. **Differential Analysis**

WALTER KRISTOF. **Statistical Method in Judging Profile Tests**

WILHELM ARNOLD. **New Experiments with the Pauli Test**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per issue.

Subscription: DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN

Brentanoweg 10, Germany